Efficient user clustering, receive antenna selection, and power allocation algorithms for massive MIMO-NOMA systems

Article (Published Version)

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 5, 2019, accepted February 22, 2019, date of publication February 28, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902331

Efficient User Clustering, Receive Antenna Selection, and Power Allocation Algorithms for Massive MIMO-NOMA Systems

WALID A. AL-HUSSAIBI[1], (Senior Member, IEEE), AND FALAH H. ALI[2], (Senior Member, IEEE)
[1]Department of Electrical Techniques, Basrah Technical Institute, Southern Technical University, Basrah 4200, Iraq
[2]Communication Research Group, School of Engineering and Informatics, University of Sussex, Brighton BN19QT, U.K.

Corresponding authors: Walid A. Al-Hussaibi (alhussaibi@stu.edu.iq) and Falah H. Ali (f.h.ali@sussex.ac.uk)

ABSTRACT Massive multiple-input multiple-output (MIMO) and nonorthogonal multiple access (NOMA)-based technologies are considered as essential parts in the 5G systems to fulfill the escalating demands of higher connectivity and data rates for emerging wireless applications. In this paper, a new approach of massive MIMO-NOMA with receive antenna selection (RAS) is considered for the uplink channel to significantly increase the number of connected devices and overall sum rate capacity with improved user-fairness and less complexity. The proposed scheme is designed from two multiuser MIMO (MU-MIMO) clusters, based on the available number of radio frequency chains (RFCs) at the base station and channel conditions, followed by power-domain NOMA for the simultaneous signal transmission. We derive the sum rate and capacity region expressions for MIMO-NOMA with RAS over Rayleigh fading channels. Then, an optimal and three highly efficient sub-optimal dynamic user clustering, RAS, and power allocation algorithms are proposed for sum rate maximization under received power constraints and minimum rate requirements of the allowed users. The effectiveness of designed algorithms is verified through extensive analysis and numerical simulations compared to the reference MU-MIMO and MIMO-NOMA systems. The achieved results show a substantial increase in connectivity, up to two-fold for the accessible number of RFCs, and overall sum rate capacity while satisfying the minimum users' rates. Besides, important tradeoffs can be realized between system performances, hardware and computational complexities, and desired user-fairness in terms of serving more users with equal/unequal rates.

INDEX TERMS Massive MIMO-NOMA, massive connectivity, user clustering, antenna selection, power allocation, channel capacity, capacity region, user fairness.

I. INTRODUCTION

By 2020, the emergence of fifth generation (5G) mobile systems will be reality to counter the rapid explosion of global data traffic, driven mainly by the massive use of smartphones, laptops, and other smart devices/machines to get advantages of various new wireless services. Mobile Internet and massive machine-type communications (mMTC), also known as the Internet of things (IoT), are examples of such important applications that poses critical requirements for 5G cellular networks such as massive connectivity, fiber-like data rate transmission, ultra-reliable low-latency communications, wider coverage, improved user-fairness, and flexible multiple access (MA) schemes [1]–[4]. Therefore, several promising techniques have been proposed by the research and industrial communities to meet these critical challenges. In particular and due to the limited wireless spectrum and power resources, it is highly envisioned that massive (or large scale) multiple-input multiple-output (MIMO) [3], [4] and nonorthogonal multiple access (NOMA) [1], [2] will be the key components for 5G and beyond. The early versions of these technologies have been already adopted in Releases 13 and 14 of Third Generation Partnership Project (3GPP) Long Term Evolution Advanced Pro (LTE-A Pro) [5], [6]. Furthermore, enhanced specifications are incorporated in the first specification of 5G New Radio (NR) standard under Release 15 of non-standalone and standalone operations [3], [4], [7].

A. BACKGROUND

Massive MIMO is achieved by equipping the base station (BS) with hundreds to thousands of antennas associated

The associate editor coordinating the review of this manuscript and approving it for publication was Jie Xu.

with radio frequency chains (RFCs) to enhance the spatial degree of freedom (DoF) and array gain considerably. This approach allows simultaneous transmission of tens to hundreds of mobile user equipments (UEs) without consuming extra power and subdivision in the scarce resources of time, frequency, and codes [8], [9]. It has been confirmed, through extensive analysis and results, that massive MIMO systems enable significant improvement in spectral efficiency, reliability, and link latency [10]–[15]. At present, Releases 13 and 14 of LTE-A Pro standards support up to 16 and 32 antennas at the BS, respectively with an increased number of up to 8 co-scheduled UEs. In addition, Release 15 of the 5G NR expands the BS with 64 to hundreds of antennas [4], [5]. So far, several designs have shown the possibility of implementing large number of antennas however at the cost of high complexity and power consumption owing to the need for same number of associated RFCs [10], [11], [13]. Therefore, antenna selection (AS) has been considered as an effective solution to reduce the impact of these critical problems [16], [17] and becomes an essential part of multiantenna systems [10]–[14], [18]–[21]. This technique has been investigated over different channel conditions by exploiting the spatial diversity, represented by the difference between higher number of antennas and available RFCs, and demonstrated significant performance gains [11], [19].

On the other hand, power-domain NOMA schemes based on superposition coding (SC) at the transmit side and successive interference cancellation (SIC) at the receiver has been recognized lately as a promising technology to simultaneously serve multiple users with different channel conditions at the same time, frequency, code, and spatial DoF [22]–[25]. Compared with the most efficient orthogonal MA (OMA) schemes represented by orthogonal frequency division MA (OFDMA) in 4G networks, it enables extra number of connected UEs, higher spectral efficiency, and diverse quality-of-service (QoS) [2]. Therefore, NOMA has been considered in 3GPP LTE-A Pro for the downlink under multiuser superposition transmission (MUST) scheme [1]. Moreover, it is envisioned that 5G NR will support the application of uplink NOMA to provide large-scale connectivity for mMTC and other applications [7]. However, NOMA has many research challenges that need careful investigations such as optimal/sub-optimal resource allocations, dynamic user clustering algorithms, low complexity SIC based receivers and multiuser detection (MUD) methods, and in-depth capacity analysis when combined with MIMO systems [2], [26]–[30].

The integration of NOMA concepts in MIMO systems can support extra UEs and enhance the performance gains compared with the existing MIMO-OMA schemes due to the additional DoFs. Therefore, MIMO-NOMA becomes a hot research topic and attracted high attention for both downlink and uplink channels [31]–[38]. To achieve the promised gains, the allowed users are usually grouped in clusters based on their propagation channel conditions, and different strategies have been proposed for power allocation [29], [39], [40], cluster/group formation [35], [38], [41], and signal precoding and detection [31], [33], [34], [37]. Massive MIMO-NOMA designs are also considered for different targets such as achievable rate enhancement [1], multiresolution multicast services [29], extending the user capacity with low receiver complexity [38], maintaining performance gains with partial channel state information (CSI) [42], robust vehicular communications with spatial modulation [43], resource allocations when combined with relaying schemes [44], joint beamforming and power allocation [45], and security for opportunistic multicast transmissions [46]. Nevertheless, more research activities are required for critical system designs that take into account the aforesaid challenges for massive MIMO-NOMA systems.

B. AIMS AND CONTRIBUTIONS

To provide effective solutions for the requirements of 5G systems without excess in spectrum and power resources, this work aims to maximize the connectivity and sum rate capacity of massive MIMO-NOMA uplink channel with enhanced user-fairness and less system complexity. User-fairness can be viewed in terms of equal rate distribution among all connected UEs regardless of their channel conditions or unequal users' rates by serving strong channel users with maximum rates and satisfying the minimum rate target (QoS) of weak channel users [22], [23], [35]. On the other hand, system complexity represents a key issue to minimize the consumed power towards safe and green communications [11], [12], [47]. For instance, the hardware complexity, in terms of utilized RFCs, is responsible for 50-80% of the total power consumption at the BS [12], while the essential computational efforts consume about 20% of the rest power [47]. Thus, reducing the number of RFCs and computational burden represent crucial factors for given user connectivity and sum rate capacity targets.

In this paper, a new approach of massive MIMO-NOMA uplink with receive AS (RAS) is considered based on the adopted design in our recent works [14], [38] without use of signal alignment [31], channel coding [33], [37], multicarrier transmission [36], or signal precoding [34]. The allowed users are divided into two multiuser MIMO (MU-MIMO) clusters based on the accessible number of RFCs and channel conditions, denoted as high power cluster (HPC) and low power cluster (LPC). Simultaneous signal transmission from both clusters is maintained by employing power-domain NOMA with efficient inter-cluster and intra-cluster power allocation strategies. At the BS, the received superimposed signals associated with RAS are estimated reliably using two low complexity linear MUD stages and SIC process.

The main contributions of this paper are summarized as follows:

- We derive the sum rate capacity for the uplink massive MIMO-NOMA channel with RAS over Rayleigh fading environment. In addition, we provide the capacity region expressions for the achievable clusters' sum rates and

present the optimal operating point that maximize the overall sum rate capacity. This important point identifies the optimal tradeoffs between the overall sum rate, clusters' sum rates, users' rates, and user overloading (i.e. size of LPC).
- For overall sum rate maximization under received power restrictions and minimum rate requests of HPC and LPC users, we present an optimal algorithm for the dynamic user clustering, RAS, and power allocation based on the joint exhaustive search. To mitigate the ultra-high computational efforts, we propose three highly efficient sub-optimal algorithms by splitting the joint problem into low complexity components. Detailed complexity analysis of all designed algorithms is presented.
- Since the 5G NR supports backward compatibility with the 3GPP LTE systems [4], [7], we demonstrate the effectiveness of the proposed algorithms through extensive analysis and numerical simulations for different moderate and large scale system scenarios compared with the conventional MU-MIMO and MIMO-NOMA schemes in [48] and [49]. The achieved outcomes validate massive increase in connected users, up to double number of utilized RFCs, and higher overall sum rate capacity. Moreover, vital tradeoffs are demonstrated between achieved connectivity, overall sum rate, clusters' sum rates, required number of RFCs, computational complexity, and user-fairness of equal/unequal rate distribution.

The rest of this paper is organized as follows: In Section II, the system design of massive MIMO-NOMA is described. Section III presents the sum rate and capacity region analysis of the considered system. Section IV deals with the dynamic user clustering, RAS, and power allocation by providing the problem formulation, proposed algorithms, and related complexity analysis. The conducted results are shown in Section V. Finally, Section VI concludes the paper.

Notations: Bold-face uppercase and lowercase letters denote matrices and vectors, respectively. Plain lowercase letters stand for scalars. $\mathcal{C}^{m\times u}$ denotes complex $m \times u$ matrix while $\mathcal{R}^{m\times u}$is for real $m\times u$ matrix. Superscripts $[.]^*$, $[.]^{\mathrm{H}}$, $[.]^{\mathrm{T}}$ and $[.]^{\dagger}$ stand for complex conjugate, conjugate transposition, transposition, and pseudoinverse, respectively. $E[.]$ stands for the expectation operator. $\mathbf{I}_m$ is $m \times m$ identity matrix and $\|.\|$ stands for the Euclidean vector norm. $|.|$ denote the determinant for matrices and magnitude for vectors.

II. SYSTEM DESIGN OF MASSIVE MIMO-NOMA

A. SYSTEM MODEL

Consider an uplink massive MIMO-NOMA scenario in a single cell cellular system of K randomly deployed users communicating simultaneously over flat Rayleigh fading channel with one common BS. Each mobile UE has a single-antenna while the BS which is equipped with a large array of M antennas and M_s RFCs employs RAS to select the best subset of antennas ($M_s \leq M$) based on their channel conditions. As in [29], [31]–[35], [37], [41], and [43], we assume perfect synchronization and CSI at the receiver with fading rate much less than the data rate (i.e. slowly varying) to isolate the impact of these parameters and show the actual gain of proposed approach. It should be noted that imperfect CSI and synchronization between the users are important practical issues [18], [21] however beyond the scope of this paper.

In the context of spectrally efficient wireless systems, mobile users of strong channel gains have the priority of the accessible communication links in contrast to those of weak channel conditions. On the other hand, a balance between spectral efficiency and fairness in distributing the system resources among connected UEs should be maintained to fulfill the requirements of next-generation networks [31], [35]. Motivated by these facts, user partitioning is considered in this work by dividing the allowed users dynamically into two MU-MIMO clusters based on their received signal power namely, HPC of strong channel users and LPC of weak channel users. In this scenario, power-domain NOMA is performed for signal transmission of HPC and LPC by employing power control under total received power constraint $\mathcal{P}$ during every transmission time interval. At the BS, a low complexity layered MUD is used for the received superimposed signals. It consists of two stages of linear MUD to estimate HPC signals first considering the interference from LPC as a background noise, followed by SIC to remove the contribution of HPC from received signal vector. The second stage of MUD will be used for estimating LPC signals.

The model of the received superimposed signal vector from HPC and LPC at M receive antennas $\mathbf{r} \in \mathcal{C}^{M\times 1}$ is given as

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{h}_k \sqrt{p}_k s_k + \mathbf{n} \tag{1}$$

where $\mathbf{h}_k \in \mathcal{C}^{M\times 1}$ is the composite channel vector of user k whose entries h_{mk} represent the complex gains between user k and m^{th} receive antenna due to large scale path loss and small scale fading, s_k denotes the transmitted symbol of user k with $E\left[s_k s_k^*\right] = 1$, p_k denotes the transmitted power of user k subject to maximum power constraint that the UE can handle and/or the spectrum regulations allow, and $\mathbf{n} \in \mathcal{C}^{M\times 1}$ is i.i.d complex additive white Gaussian noise (AWGN) vector with elements having zero mean and variance σ_n^2. The channel vector $\mathbf{h}_k$ can be represented as [31]

$$\mathbf{h}_k = \frac{\mathbf{g}_k}{\sqrt{L(d_k)}} \tag{2}$$

$$L(d_k) = \begin{cases} d_k^{\zeta}, & d_k > d_0 \\ d_0^{\zeta}, & d_k \leq d_0 \end{cases} \tag{3}$$

where $\mathbf{g}_k = [g_{1k} \cdots g_{Mk}]^{\mathrm{T}} \in \mathcal{C}^{M\times 1}$ is the Rayleigh fading channel vector of user k whose entries g_{mk} are zero mean unit variance complex Gaussian coefficient between user k and m^{th} receive antenna, $L(d_k)$ denotes the path loss of user k located at a distance d_k from the BS and assumed to be the same at each receive antenna, d_0 is the reference distance according to cell size, and ζ denotes the path loss exponent.

B. DYNAMIC USER CLUSTERING

In view of the fact that each of the two considered linear MUD stages has total DoFs equal to M_s RFCs, the allowed number of connected UEs (streams) is upper bounded by $K \leq 2M_s$. User clustering in HPC is formed from $T = M_s$ users of highest received powers to satisfy the channel rank condition and preserve the maximum connectivity and sum rate of the generic MU-MIMO with linear MUD. On the other hand, LPC is configured from the rest of users $U = (K - T) \leq M_s$ of lowest powers to attain the user-fairness with acceptable interference level to HPC users. Note that the additional U users of weak channel conditions are commonly terminated in the basic MU-MIMO schemes. Therefore, the range of supported UEs in the considered massive MIMO-NOMA system is bounded by $M_s < K \leq 2M_s$.

Practically, dynamic user clustering can be achieved based on the channel path loss for each user $L(d_k); k = 1, \ldots, K$ which is inversely proportional to the average received signal power. By adopting this strategy, users near to the BS are highly probable to be included in HPC due to strong channel conditions in contrast to those located at far distances, which results in an improved spectral efficiency and user-fairness. Consequently, for full hardware complexity system of $M_s = M$, the basic design criterion for cluster formation can be given as follows:

1) Calculate the channel path losses $L(d_k)$; $k = 1, \ldots, K$.
2) Define $\Psi = [1, 2, \ldots, K]$ as the set of all active users, sorted according to their path loss parameters in ascending order, i.e. $L(d_1) < L(d_2) < \ldots < L(d_K)$.
3) Construct the set of HPC as $\Phi = [1, \ldots, T]$ from the first T elements in Ψ that represent strong users.
4) Construct the set of LPC as $\Theta = [1, \ldots, U]$ from the rest U elements in Ψ that represent weak users.
5) Repeat steps 1 to 4 whenever users' locations changed (i.e. path losses) to update the sets Φ and Θ.

Considering the designed HPC and LPC for full complexity system, the signal model (1) can be rewritten as

$$\begin{aligned}\mathbf{r} &= \underbrace{\sum_{i=1, i\in\Phi}^{T} \mathbf{h}_i\sqrt{p_i}s_i}_{\text{HPC}} + \underbrace{\sum_{j=1, j\in\Theta}^{U} \mathbf{h}_j\sqrt{p_j}s_j}_{\text{LPC}} + \mathbf{n} \\ &= \mathbf{H}_H\mathbf{x}_H + \mathbf{H}_L\mathbf{x}_L + \mathbf{n} \end{aligned} \tag{4}$$

where $\mathbf{H}_H = [\mathbf{h}_1 \cdots \mathbf{h}_T] \in \mathcal{C}^{M\times T}$ and $\mathbf{H}_L = [\mathbf{h}_1 \cdots \mathbf{h}_U] \in \mathcal{C}^{M\times U}$ are the subchannels of HPC and LPC, respectively, $\mathbf{x}_H = \left[\sqrt{p}_1 s_1 \cdots \sqrt{p}_T s_T\right]^{\mathrm{T}} \in \mathcal{C}^{T\times 1}$ is the transmitted signal vector from HPC, and $\mathbf{x}_L = \left[\sqrt{p}_1 s_1 \cdots \sqrt{p}_U s_U\right]^{\mathrm{T}} \in \mathcal{C}^{U\times 1}$ is the transmitted signal vector from LPC.

C. POWER ALLOCATION

Based on the basic implementation principles of power-domain NOMA [22], [31], [41], the power difference between received signals from designed HPC and LPC is essential to manage the inter-cluster interference and perform efficient SIC process at the receiver. In addition, total average received power constraint $\mathcal{P}$ represents a crucial design criterion to reduce the power consumption at the UEs (i.e. prolongs the batteries lifetime) and minimize the intra-cell as well as inter-cell interference. Therefore, the average received powers at full complexity BS from HPC users ($\mathcal{P}_H$) and LPC users ($\mathcal{P}_L$) are specified during every transmission time period as $\mathcal{P}_H + \mathcal{P}_L = \mathcal{P}$, and maintained through the inter-cluster power allocation policy to satisfy target user rates as

$$\mathcal{P}_H = \beta_H \mathcal{P} = \sum_{i=1, i\in\Phi}^{T} p_i \frac{\alpha}{L(d_i)} \tag{5}$$

$$\mathcal{P}_L = \beta_L \mathcal{P} = \sum_{j=1, j\in\Theta}^{U} p_j \frac{\alpha}{L(d_j)} \tag{6}$$

where the factor α is used to insure that the transmitted power from each UE do not exceed its maximum rating, β_H and β_L are dynamic power allocation coefficients for HPC and LPC, respectively with $\beta_H + \beta_L = 1$ and $\beta_H > \beta_L > 0$.

For users within each cluster, statistics-aware intra-cluster power allocation is used to compensate the disparities between users' signal attenuations. This strategy has the advantage of allowing uniform user performance within each cluster due to equal effective channel gains for supported UEs as $\left\{p_i\alpha/L(d_i) = \mathcal{P}_H/T\right\}_{i=1}^{T}$ and $\left\{p_j\alpha/L(d_j) = \mathcal{P}_L/U\right\}_{j=1}^{U}$ for HPC and LPC, respectively. Consequently, the allocated transmit power for each user can be expressed in terms of the associated path loss and premeditated cluster's parameters as

$$p_i = \frac{\beta_H \mathcal{P}}{\alpha T} L(d_i); \quad i \in \Phi,\ i = 1, \ldots, T \tag{7}$$

$$p_j = \frac{\beta_L \mathcal{P}}{\alpha U} L(d_j); \quad j \in \Theta,\ j = 1, \ldots, U. \tag{8}$$

D. RAS TECHNIQUE

In MIMO systems, implementing more RFCs to support massive number of UEs is impractical in terms of hardware requirements, consumed power, and increased receiver size [17]. Therefore, AS is usually used to capture most of the massive MIMO gains when the number of utilized antennas is higher than available RFCs by utilizing inexpensive RF switches and digital signal processing circuitry [13]. In the literature, different AS algorithms have been proposed to select the best subset of antennas, mostly based on highest received power [11], [38] or capacity maximization [12]–[14], [19]. The latter approach is known to offer optimal/near-optimal performance compared with the former at the cost of exhaustive search requirements for antenna subset selection (grows exponentially with M), which is computationally prohibitive [10]. But, the presented sum rate capacity results in [11] for power based selection (PBS) over real propagation environment demonstrated very close performance to that of near-optimal capacity based selection (CBS). Based on this finding, a simplified binary switching has been used for AS in [13] to maximize the channel capacity and shows significant complexity reduction compared with

the full switching scheme at the cost of small performance loss.

In this work, RAS is utilized to achieve massive increase in sum rate capacity of designed system with affordable complexity. For this purpose, PBS and CBS techniques are proposed in Section 4 to select the best subset of M_s from M receive antennas as $\mathcal{s}_l \in \mathcal{S}$, where$\mathcal{S} = \{\mathcal{s}_1, \ldots, \mathcal{s}_l, \ldots, \mathcal{s}_{|\mathcal{S}|}\}$ represent the set of all potential subsets whose cardinality is $|\mathcal{S}| = \binom{M}{M_s} = \frac{M!}{M_s!(M-M_s)!}$. The selection process is based on the overall channel matrix $\mathbf{H} \in \mathcal{C}^{M\times K}$ of considered system represented as

$$\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_k \cdots \mathbf{h}_K] = [\mathbf{b}_1, \cdots, \mathbf{b}_m, \cdots, \mathbf{b}_M]^{\mathrm{T}} \quad (9)$$

where $\mathbf{h}_k \in \mathcal{C}^{M\times 1}$ is the k^{th} column corresponding to user k and $\mathbf{b}_m \in \mathcal{C}^{1\times K}$ is the m^{th} row corresponding to m^{th} receive antenna.

From (4), the received signal vector associated with user clustering and RAS can be written as

$$\begin{aligned}\check{\mathbf{r}} &= \sum_{i=1,i\in\Phi}^{T} \check{\mathbf{h}}_i\sqrt{p}_i s_i + \sum_{j=1,j\in\Theta}^{U} \check{\mathbf{h}}_j\sqrt{p}_j s_j + \check{\mathbf{n}} \\ &= \check{\mathbf{H}}_H\mathbf{x}_H + \check{\mathbf{H}}_L\mathbf{x}_L + \check{\mathbf{n}} \end{aligned} \quad (10)$$

where $\check{\mathbf{r}} \in \mathcal{C}^{M_s\times 1}$, $\check{\mathbf{h}}_k \in \mathcal{C}^{M_s\times 1}$, and $\check{\mathbf{n}} \in \mathcal{C}^{M_s\times 1}$ denote received signal, k^{th} user channel, and noise vectors after selection, respectively, $\check{\mathbf{H}}_H \in \mathcal{C}^{M_s\times T}$ and $\check{\mathbf{H}}_L \in \mathcal{C}^{M_s\times U}$ are HPC and LPC channels associated with RAS, respectively and can be used to characterize the overall channel matrix after user clustering and RAS as $\check{\mathbf{H}} = \left[\check{\mathbf{H}}_H \check{\mathbf{H}}_L\right] \in \mathcal{C}^{M_s\times K}$.

III. SUM RATE AND CAPACITY REGION ANALYSIS

A. SUM RATE CAPACITY

The sum rate capacity of the proposed massive MIMO-NOMA without RAS is given in terms of the achievable sum rates of HPC (R_H) and LPC (R_L) as

$$R_{sum} = R_H + R_L = \sum_{i=1,i\in\Phi}^{T} R_i + \sum_{j=1,j\in\Theta}^{U} R_j \quad (11)$$

where R_i and R_j are the achievable rates of i^{th} user and j^{th} user within HPC and LPC, respectively according to their allocated transmit powers and bounded by $R_i \geq R_j \geq R_0$; $\forall i, j$, where R_0 stands for the considered minimum user rate constraint to warrant the QoS requirements. It should be noted that users in each cluster will have equal average rate distribution since they have equal effective channel gains.

Based on the capacity of uplink MU-MIMO channel [48], [49], the sum rate capacity of signal model (4) can be given for constant channel realization as

$$R_{sum} \leq \log_2\left|\mathbf{I}_M + \frac{1}{\sigma_n^2}\left(\sum_{\substack{i=1,\\ i\in\Phi}}^{T}\mathbf{h}_i p_i \mathbf{h}_i^{\mathrm{H}} + \sum_{\substack{j=1,\\ j\in\Theta}}^{U}\mathbf{h}_j p_j \mathbf{h}_j^{\mathrm{H}}\right)\right|. \quad (12)$$

Using (7) and (8) of the allocated users' powers, the above equation can be written in terms of power allocation coefficients β_H and β_L as

$$\begin{aligned} R_{sum} \leq \log_2\Bigg|\mathbf{I}_M + \frac{\gamma}{\alpha}\Bigg(&\frac{\beta_H}{T}\sum_{i=1,i\in\Phi}^{T} L\left(d_i\right)\mathbf{h}_i\mathbf{h}_i^{\mathrm{H}} \\ &+\frac{\beta_L}{U}\sum_{j=1,j\in\Theta}^{U} L\left(d_j\right)\mathbf{h}_j\mathbf{h}_j^{\mathrm{H}}\Bigg)\Bigg| \end{aligned} \quad (13)$$

$$R_{sum} \leq \log_2\left|\mathbf{I}_M + \frac{\gamma}{\alpha}\left(\frac{\beta_H}{T}\mathbf{D}_H + \frac{\beta_L}{U}\mathbf{D}_L\right)\right| \quad (14)$$

where $\gamma = \mathcal{P}/\sigma_n^2$ is the average SNR at each receive antenna while the matrices $\mathbf{D}_H = \sum_{i=1,i\in\Phi}^{T} L\left(d_i\right)\mathbf{h}_i\mathbf{h}_i^{\mathrm{H}} \in \mathcal{C}^{M\times M}$ and $\mathbf{D}_L = \sum_{j=1,j\in\Theta}^{U} L\left(d_j\right)\mathbf{h}_j\mathbf{h}_j^{\mathrm{H}} \in \mathcal{C}^{M\times M}$ are related to HPC and LPC, respectively.

With RAS, the sum rate R_{sum}^{s} can be maximized as

$$R_{sum}^{s} \leq \max_{\substack{\mathcal{s}_l\in\mathcal{S}\\ l=1,\ldots,|\mathcal{S}|}}\left\{\log_2\left|\mathbf{I}_{M_s} + \frac{\gamma}{\alpha}\left(\frac{\beta_H}{T}\check{\mathbf{D}}_H + \frac{\beta_L}{U}\check{\mathbf{D}}_L\right)\right|\right\} \quad (15)$$

where the maximization process is performed over subset $\mathcal{s}_l \in S; l = 1, \ldots, |\mathcal{S}|$, and the matrices $\check{\mathbf{D}}_H \in \mathcal{C}^{M_s\times M_s}$ and $\check{\mathbf{D}}_L \in \mathcal{C}^{M_s\times M_s}$ associated with RAS can be found as

$$\check{\mathbf{D}}_H = \sum_{i=1,i\in\Phi}^{T} L\left(d_i\right)\check{\mathbf{h}}_i\check{\mathbf{h}}_i^{\mathrm{H}} \quad (16)$$

$$\check{\mathbf{D}}_L = \sum_{j=1,j\in\Theta}^{U} L\left(d_j\right)\check{\mathbf{h}}_j\check{\mathbf{h}}_j^{\mathrm{H}} \quad (17)$$

Therefore, the resulting ergodic sum rate over randomly varying channel realizations is given by

$$\begin{aligned}&\mathbb{E}\left[R_{sum}^{s}\right] \\ &\leq \mathbb{E}\left[\max_{\substack{\mathcal{s}_l\in\mathcal{S}\\ l=1,\ldots,|\mathcal{S}|}}\left\{\log_2\left|\mathbf{I}_{M_s} + \frac{\gamma}{\alpha}\left(\frac{\beta_H}{T}\check{\mathbf{D}}_H + \frac{\beta_L}{U}\check{\mathbf{D}}_L\right)\right|\right\}\right].\end{aligned} \quad (18)$$

B. CAPACITY REGION

Considering the sum rates of designed clusters, R_H and R_L, capacity region of massive MIMO-NOMA system without RAS ($M = M_s$) and over constant channel realization can be given based on the capacity expressions of MU-MIMO schemes [49] and (14) as the set of all sum rates (R_H, R_L) satisfying the following three constraints

$$R_H \leq \log_2\left|\mathbf{I}_M + \frac{\gamma\beta_H}{\alpha T}\mathbf{D}_H\right| \quad (19)$$

$$R_L \leq \log_2\left|\mathbf{I}_M + \frac{\gamma\beta_L}{\alpha U}\mathbf{D}_L\right| \quad (20)$$

$$R_{sum} = R_H + R_L \leq \log_2\left|\mathbf{I}_M + \frac{\gamma}{\alpha}\left(\frac{\beta_H}{T}\mathbf{D}_H + \frac{\beta_L}{U}\mathbf{D}_L\right)\right|. \quad (21)$$

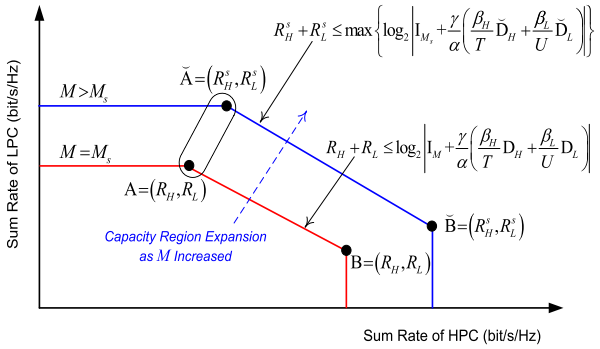

FIGURE 1. Capacity region of massive MIMO-NOMA system for $M \geq M_s$.

The capacity region is shown in Fig. 1 where the maximum sum rate points $R_H(\mathrm{B})$ for HPC and $R_L(\mathrm{A})$ for LPC can be achieved through (19) and (20), respectively as if the other cluster is absent from the system. On the other hand, constraint (21) is related to the achievable overall sum rate when users of both clusters are communicating simultaneously with the BS receiver. The corner points, $\mathrm{A} = (R_H, R_L)$ and $\mathrm{B} = (R_H, R_L)$, can be achieved by applying SIC whereas all other points on the line $\overline{\mathrm{AB}}$ can be realized through SIC with time or frequency sharing.

For the considered system, signals of HPC are designed to be decoded first with the presence of interference from LPC, followed by SIC to decode the signals of LPC. Therefore, the achievable sum rate capacity of the optimal operating point A is given as

$$R_H(\mathrm{A}) = \log_2 \left| \mathbf{I}_M + \frac{\gamma \beta_H}{\alpha T} \left[\mathbf{I}_M + \frac{\gamma \beta_L}{\alpha U} \mathbf{D}_L \right]^{-1} \mathbf{D}_H \right| \tag{22}$$

$$R_L(\mathrm{A}) = \log_2 \left| \mathbf{I}_M + \frac{\gamma \beta_L}{\alpha U} \mathbf{D}_L \right|. \tag{23}$$

Note that if the detection process is reversed by estimating signals of LPC first with HPC interference (not feasible for this system but for illustration purpose), the sum rate point B can be given as

$$R_H(\mathrm{B}) = \log_2 \left| \mathbf{I}_M + \frac{\gamma \beta_H}{\alpha T} \mathbf{D}_H \right| \tag{24}$$

$$R_L(\mathrm{B}) = \log_2 \left| \mathbf{I}_M + \frac{\gamma \beta_L}{\alpha U} \left[\mathbf{I}_M + \frac{\gamma \beta_H}{\alpha T} \mathbf{D}_H \right]^{-1} \mathbf{D}_L \right|. \tag{25}$$

The considered operating point A on the capacity region can be used to characterize the optimal tradeoff between the overall achievable sum rate, clusters' sum rates, users' rates, and user overloading U (i.e. size of LPC).

When RAS of subset $\mathcal{S}_l \in S; l = 1, \ldots, |\mathcal{S}|$ is utilized, the associated capacity region (see Fig. 1) can be demonstrated as the set of all sum rates $\left(R_H^s, R_L^s\right)$ fulfilling the following three conditions

$$R_H^s \leq \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_H}{\alpha T} \check{\mathbf{D}}_H \right| \right\} \tag{26}$$

$$R_L^s \leq \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_L}{\alpha U} \check{\mathbf{D}}_L \right| \right\} \tag{27}$$

$$R_{sum}^s = R_H^s + R_L^s \leq \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma}{\alpha} \left(\frac{\beta_H}{T} \check{\mathbf{D}}_H + \frac{\beta_L}{U} \check{\mathbf{D}}_L \right) \right| \right\}. \tag{28}$$

In this case, the operating sum rate point $\check{A} = \left(R_H^s, R_L^s\right)$ and the other corner point $\check{B} = \left(R_H^s, R_L^s\right)$ can be found as

$$R_H^s(\check{A}) = \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_H}{\alpha T} \left[\mathbf{I}_{M_s} + \frac{\gamma \beta_L}{\alpha U} \check{\mathbf{D}}_L \right]^{-1} \check{\mathbf{D}}_H \right| \right\} \tag{29}$$

$$R_L^s(\check{A}) = \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_L}{\alpha U} \check{\mathbf{D}}_L \right| \right\} \tag{30}$$

$$R_H^s(\check{B}) = \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_H}{\alpha T} \check{\mathbf{D}}_H \right| \right\} \tag{31}$$

$$R_L^s\left(\check{B}\right) = \max_{\mathcal{S}_l \in \mathcal{S}; \beta_H > \beta_L} \left\{ \log_2 \left| \mathbf{I}_{M_s} + \frac{\gamma \beta_L}{\alpha U} \left[\mathbf{I}_{M_s} + \frac{\gamma \beta_H}{\alpha T} \check{\mathbf{D}}_H \right]^{-1} \check{\mathbf{D}}_L \right| \right\}. \tag{32}$$

It should be noted that when M is reduced to M_s, the sum rates R_{sum}^s, R_H^s, and R_L^s will be identical to R_{sum}, R_H, and R_L, respectively. As a result, the corner points $\check{A}$ and $\check{B}$ on the capacity region will be shifted backwards to points A and B, respectively (i.e. $R_H^s(\check{A}) = R_H(\mathrm{A})$, $R_L^s(\check{A}) = R_L(\mathrm{A})$, $R_H^s(\check{B}) = R_H(\mathrm{B})$, and $R_L^s(\check{B}) = R_L(\mathrm{B})$).

IV. DYNAMIC USER CLUSTERING, RAS, AND POWER ALLOCATION ALGORITHMS

In this section, we propose dynamic user clustering, RAS, and power allocation algorithms for massive MIMO-NOMA system to provide the important tradeoffs between the achieved performance (R_{sum}^s, R_H^s, R_L^s, and additional users U), system complexity (number of RFCs and computational efforts), and user-fairness (equal/unequal user rates).

A. PROBLEM FORMULATION

Efficient user clustering, RAS, and power allocation techniques are developed for the considered system based on the following power and minimum rate constraints:

1) POWER CONSTRAINTS

$$\mathrm{C1}: \mathcal{P} = \sum_{k=1}^{K} p_k \frac{\alpha}{L(d_k)} = \underbrace{\sum_{i=1, i \in \Phi}^{T} p_i \frac{\alpha}{L(d_i)}}_{\mathcal{P}_H = \beta_H \mathcal{P}} + \underbrace{\sum_{j=1, j \in \Theta}^{T} p_j \frac{\alpha}{L(d_j)}}_{\mathcal{P}_L = \beta_L \mathcal{P}} \tag{33}$$

$$\text{C2}: \mathcal{P}_H - \mathcal{P}_L = (\beta_H - \beta_L)\,\mathcal{P} \geq \mathcal{P}_{dif} \tag{34}$$

where constraint C1 is related to the total received power ($\mathcal{P}$) from both of HPC and LPC users while constraint C2 ensures the essential minimum power difference $\mathcal{P}_{dif}$ between $\mathcal{P}_H$ and $\mathcal{P}_L$ to handle the inter-cluster interference and perform efficient SIC process at the receiver.

2) MINIMUM RATE CONSTRAINTS

$$\text{C3}: R_H^s = \sum_{i=1, i\in\Phi}^{T} R_i^s \geq TR_0 \tag{35}$$

$$\text{C4}: R_L^s = \sum_{j=1, j\in\Theta}^{U} R_j^s \geq UR_0 \tag{36}$$

where C3 and C4 are used to warrant the minimum sum rates for HPC and LPC, respectively. R_i^s and R_j^s denote the achievable rates of i^{th} and j^{th} users within HPC and LPC after RAS, respectively. So, the upper and lower bounds of overall sum rate, R_{sum}^s, can be given based on (28), (35) and (36) as

$$KR_0 \leq R_{sum}^s \leq \max_{\mathcal{S}_l\in\mathbb{S};\beta_H>\beta_L} \times \left\{\log_2\left|\mathbf{I}_{M_s} + \frac{\gamma}{\alpha}\left(\frac{\beta_H}{T}\check{\mathbf{D}}_H + \frac{\beta_L}{U}\check{\mathbf{D}}_L\right)\right|\right\}. \tag{37}$$

To maximize the overall sum rate for each channel realization under C1 − C4, the optimization problem can be formulated in accordance with the user clustering and operating sum rate point $\check{A} = \left(R_H^s, R_L^s\right)$ as

$$\max_{\substack{(\Phi,\Theta)_n\in\Sigma \\ \mathcal{S}_l\in\mathbb{S};\beta_H>\beta_L}} \underbrace{\log_2\left|\mathbf{I}_{M_s} + \frac{\gamma\beta_H}{\alpha T}\left[\mathbf{I}_{M_s} + \frac{\gamma\beta_L}{\alpha U}\check{\mathbf{D}}_L\right]^1\check{\mathbf{D}}_H\right|}_{R_H^s} + \underbrace{\log_2\left|\mathbf{I}_{M_s} + \frac{\gamma\beta_L}{\alpha U}\check{\mathbf{D}}_L\right|}_{R_L^s}$$

$$\text{subject to}: \text{C1} - \text{C4}. \tag{38}$$

where $(\Phi, \Theta)_n \in \Sigma$ is the n^{th} subsets of HPC and LPC users from the overall possible user clustering set $\Sigma = \left\{(\Phi, \Theta)_1, \ldots, (\Phi, \Theta)_{|\Sigma|}\right\}$ whose cardinality is related to the number T as $|\Sigma| = \binom{K}{T} = \binom{K}{M_s}$.

Note that the overall sum rate maximization (38) is mixed-integer nonlinear programming problem that requires exhaustive search for the optimal joint user clustering and RAS over all possible combinations $\binom{K}{M_s}\binom{M}{M_s}$. Besides, closed-form solution for the optimal power allocation coefficients (β_H^* and β_L^*) is very difficult to be derived due to the determinants operations, interference term in R_H^s part, and RAS process. Consequently, an additional search for the optimal power coefficients is required based on C1 and C2 over the following ranges

$$\frac{\mathcal{P} + \mathcal{P}_{dif}}{2\mathcal{P}} \leq \beta_H < 1 \tag{39}$$

$$0 < \beta_L \leq \frac{\mathcal{P} - \mathcal{P}_{dif}}{2\mathcal{P}}. \tag{40}$$

Since the range of β_L is basically smaller than that of β_H, the inequality (40) is adopted in this work to reduce the computational complexity and obtain the optimal β_L^* using Ω division steps of equal sizes $\mu = \left(\mathcal{P} - \mathcal{P}_{dif}\right)/2\mathcal{P}\Omega$ while β_H^* can be calculated as $\beta_H^* = 1 - \beta_L^*$. However, the overall complexity, $\binom{K}{M_s}\binom{M}{M_s}\Omega$, becomes significantly high and practically impossible for massive MIMO-NOMA. Therefore, we develop three low complexity sub-optimal methods in the following subsection after presenting the optimal algorithm based on the exhausted search.

B. PROPOSED ALGORITHMS

• *Optimal Joint User Clustering, RAS, and Power Allocation (OJUC-RAS-PA) Scheme* (Algorithm 1): The optimization process is performed by searching for the optimal power coefficients for all possible subset combinations from sets Σ and S. The subsets, $(\Phi, \Theta)_n \in \Sigma$ and $\mathcal{S}_l \in S$, that maximize the sum rate (38) will be selected based on optimal indices $(n, l)^*$ to find the corresponding system parameters (i.e. $\check{\mathbf{H}}_H$, $\check{\mathbf{H}}_L$, β_H^*, β_L^*, $R_H^s(\check{A})$, $R_L^s(\check{A})$, and $R_{sum}^s(\check{A})$). This operation is dynamically updated whenever the users' channel realizations $\mathbf{h}_k; k = 1, \ldots, K$ are changed. The pseudocode of this scheme is shown in Algorithm 1.

Note that Algorithm 1 is *optimal* for the considered detection technique which consists of two linear MUD stages (for HPC and LPC) and SIC process for NOMA. However, it requires ultra-high computational complexity due to exhaustive search prerequisite for the optimal solution. To mitigate this critical drawback, the following *sub-optimal* algorithms are proposed by splitting the joint problem of sum rate maximization into three separate steps of the main components (i.e. user clustering, RAS, and power allocation).

• *User Clustering, RAS, and Power Allocation (UC-RAS-PA) Scheme* (Algorithm 2): A simple user clustering technique is performed first to find Φ and Θ sets based on the users' channel path loss parameters. The channel matrix associated with user clustering is constructed as $\mathbf{H} = [\mathbf{H}_H \mathbf{H}_L]$. This operation is followed by RAS using low complexity PBS method to choose the best subset of receive antennas $\mathcal{S}_l \in S$ and construct the subchannels of HPC and LPC associated with user clustering and RAS as $\check{\mathbf{H}}_H$ and $\check{\mathbf{H}}_L$, respectively. Then, search for the optimal power coefficients (β_H^* and β_L^*) that maximize (38) is adopted, and the corresponding system parameters (i.e. $R_H^s(\check{A})$, $R_L^s(\check{A})$, and $R_{sum}^s(\check{A})$) are obtained accordingly. The procedure of this technique is repeated once the users' channels are changed. The pseudocode of this scheme is shown in Algorithm 2. It demonstrate the lowest overall computational efforts as can be seen in the next subsection.

• *User Clustering, Power Allocation, and Power Based RAS (UC-PA-PBRAS) Scheme* (Algorithm 3): User clustering is executed first to locate Φ and Θ sets based on the users' path losses, and the associated channel matrix is constructed accordingly as $\mathbf{H} = [\mathbf{H}_H \mathbf{H}_L]$. This operation is followed by searching for the optimal power coefficients (β_H^* and β_L^*) that

maximize R_H using (22) and R_L using (23). Subsequently, a simple PBRAS method is carried out to find the best subset of receive antennas $s_l \in S$ and construct the subchannels of HPC and LPC associated with user clustering and RAS as $\check{\mathbf{H}}_H$ and $\check{\mathbf{H}}_L$, respectively. The corresponding system parameters can be found accordingly as: $R_H^s\left(\check{A}\right)$ using (29), $R_L^s\left(\check{A}\right)$ using (30), and $R_{sum}^s\left(\check{A}\right) = R_H^s\left(\check{A}\right) + R_L^s\left(\check{A}\right)$. The procedure of this algorithm is repeated once the users' channels are changed. The pseudocode of this scheme is shown in Algorithm 3.

• *User Clustering, Power Allocation, and Capacity Based RAS (UC-PA-CBRAS) Scheme* (Algorithm 4): The steps of user clustering and power allocation are similar to that of Algorithm 3. But, CBRAS (of higher complexity than PBRAS) is performed by calculating the clusters' sum rates, $R_{H(l)}^s$ using (29) and $R_{L(l)}^s$ using (30), for each antenna subset $\mathcal{s}_l \in S$. The antenna subset that maximize the sum rate $R_{sum(l)}^s = R_{H(l)}^s + R_{L(l)}^s$ will be selected as the optimal with index l^* to find the corresponding system parameters (i.e. $\check{\mathbf{H}}_H$, $\check{\mathbf{H}}_L$, $R_H^s(\check{A})$, $R_L^s(\check{A})$, and $R_{sum}^s(\check{A})$). The procedure of this method is repeated whenever the users' channels are changed. The pseudocode of this scheme is shown in Algorithm 4.

C. COMPLEXITY ANALYSIS

The complexity of proposed algorithms is presented in this subsection considering the dominant efforts of complex-valued multiplications. For this purpose, it should be noted that each multiplication process of any two matrices, $\mathbf{A} \in \mathcal{C}^{N\times N}$ and $\mathbf{B} \in \mathcal{C}^{N\times N}$, involves N^3 calculations while the inverse of a matrix $\mathbf{A}$ requires Eigenvalue decomposition of $N^3/6$ computations [38]. Besides, to find each of $\check{\mathbf{D}}_H$ and $\check{\mathbf{D}}_L$ using (16) and (17), respectively we need a number of M_s^2 calculations whereas M^2 is required for each of $\mathbf{D}_H$ and $\mathbf{D}_L$ based on (13) and (14). In Table 1, summary of the total multiplications essential for each main item in Algorithms 1-4 is presented. These results are used to find the total computational efforts and complexity order for each algorithm as shown in Table 2.

TABLE 1. Summary of the total complex-valued multiplications for each main item in Algorithms 1-4.

Main Items	Total Complex-Valued Multiplications
R_H^s using (29) or (38)	$7M_s^3/6 + 2M_s^2$
R_L^s using (30) or (38)	M_s^2
R_H using (22)	$7M^3/6 + 2M^2$
R_L using (23)	M^2
$\|\mathbf{b}_l\|^2;\ l = 1, \ldots, M$	K^2M

As can be seen from Table 2, OJUC-RAS-PA scheme involves highest computational efforts of $\mathcal{O}\left(M_s^3\binom{K}{M_s}\binom{M}{M_s}\Omega\right)$ due to exhaustive search for the optimal solution whereas UC-RAS-PA can achieve the lowest complexity of $\mathcal{O}\left(K^2M + M_s^3\Omega\right)$. It should be noted that the number M is usually much larger than M_s in massive MIMO-NOMA systems and thus has more impact on the complexity of proposed algorithms. Therefore, a significant complexity reduction can be achieved when RAS is performed before the power allocation procedure. For instance when $M = 100$, $M_s = 10$, $K = 20$, and $\Omega = 10$, OJUC-RAS-PA requires ultra-high complexity of about 3.198×10^{22} multiplications compared to 50, 000 for UC-RAS-PA algorithm. When the power allocation is executed before RAS, the complexity order depends on the utilized RAS technique as 10, 041, 000 for UC-PA-PBRAS compared to 1.73×10^{16} for UC-PA-CBRAS scheme.

TABLE 2. Complexity comparison of the proposed algorithms.

Algorithm	Computational Efforts	Complexity Order
Algorithm 1	$(7M_s^3/6 + 2M_s^2)\binom{K}{M_s}\binom{M}{M_s}\Omega$	$\mathcal{O}\left(M_s^3\binom{K}{M_s}\binom{M}{M_s}\Omega\right)$
Algorithm 2	$K^2M + (7M_s^3/6 + 2M_s^2)\Omega$	$\mathcal{O}(K^2M + M_s^3\Omega)$
Algorithm 3	$K^2M + 7M_s^3/6 + 2M_s^2 + (7M^3/6 + 2M^2)\Omega$	$\mathcal{O}(K^2M + M_s^3 + M^3\Omega)$
Algorithm 4	$(7M^3/6 + 2M^2)\Omega + (7M_s^3/6 + 2M_s^2)\binom{M}{M_s}$	$\mathcal{O}\left(M^3\Omega + M_s^3\binom{M}{M_s}\right)$

V. NUMERICAL RESULTS

In this section, numerical results of the achieved overall sum rate, clusters' sum rates, user rate, and capacity region are presented in bit/s/Hz using Monte Carlo simulations to demonstrate the effectiveness of proposed algorithms for massive MIMO-NOMA towards 5G requirements. Using different moderate and large scale system scenarios and for notational convenience,$K \times M/M_s$ denotes MIMO-NOMA scheme of K users, M receive antennas, and M_s RFCs. In addition, Algorithms 1, 2, 3, and 4 are represented in the legends of figures as a1, a2, a3, and a4, respectively. For fair comparisons and validation of results, we consider the reference MU-MIMO and MIMO-NOMA systems in [48] and [49]. The MU-MIMO of $T = M_s$ single-antenna users (maximum connectivity) and PBS is represented as $T \times M/M_s$. For the reference MIMO-NOMA with PBS, two users with total number of transmit antennas K ($T = M_s$ for strong user and $U = K - T$ for weak user) are considered, and represented as $T, U, M/M_s$. All presented results are averaged over 10^4 channel realizations and all schemes under investigation are assumed to have same total average received power of $\mathcal{P} = 1$. The adopted simulation parameters for the uplink MIMO-NOMA systems in a single cell cellular network are: inter-site distance of 500m (i.e. the distance between the BS and cell-edge); $d_0 = 50$m; $\zeta = 3.8$; $\alpha = M_s$; $\mathcal{P}_{dif} = 0.6\mathcal{P}$; $\Omega = 20$; and $R_0 = 0.1$ bit/s/Hz.

Algorithm 1 OJUC-RAS-PA Scheme

Input: $K, M, M_s, \mathcal{P}, \mathcal{P}_{dif}, \sigma_n^2, \Omega, R_0, \alpha$, and $\mathbf{h}_k; k = 1, \ldots, K$.

1: Define $\Psi = [1, \ldots, K]$ as the set of all active users, $\Phi = [1, \ldots, T]$ as a set of HPC users, $\Theta = [1, \ldots, U]$ as a set of LPC users, and $\Upsilon = [1, \ldots, M]$ as the set of all receive antennas.
2: Use Ψ to construct the set of all possible subset combinations of HPC and LPC users as $\Sigma = \left\{(\Phi, \Theta)_1, \ldots, (\Phi, \Theta)_{|\Sigma|}\right\}$.
3: Use Υ to construct the set of all potential selected receive antenna subsets $S = \left\{\mathcal{s}_1, \ldots, \mathcal{s}_{|S|}\right\}$.
4: Find $|\Sigma| = K!/M_s!\,(K - M_s)!$, $|S| = M!/M_s!\,(M - M_s)!$, and $\mu = \left(\mathcal{P} - \mathcal{P}_{dif}\right)/2\mathcal{P}\Omega$.
5: **for** $n = 1$ *to* $|\Sigma|$ **do**
6: **for** $l = 1$ *to* $|S|$ **do**
7: Set $\beta_L = 0$ and $R^s_{sum(n,l)} = 0$.
8: Construct $\check{\mathbf{H}}_{H(n,l)}$ and $\check{\mathbf{H}}_{L(n,l)}$ according to $(\Phi, \Theta)_n$ and $\mathcal{s}_l$.
9: **for** $q = 1$ *to* Ω **do**
10: Update $\beta_L = \beta_L + \mu$.
11: Find: $\beta_H = 1 - \beta_L$ and both terms R^s_H and R^s_L in (38).
12: **if** $\left(R^s_H \geq TR_0\right)$ AND $\left(R^s_L \geq UR_0\right)$ AND $\left(R^s_H + R^s_L > R^s_{sum(n,l)}\right)$ **then**
13: Update $R^s_{sum(n,l)} = R^s_H + R^s_L$, $R^s_{H(n,l)} = R^s_H$, $R^s_{L(n,l)} = R^s_L$, and $\beta_{L(n,l)} = \beta_L$.
14: **end if**
15: **end for**
16: **end for**
17: **end for**
18: Choose the indices that satisfy the optimization problem (38):

$$(n, l)^* = \arg \max_{\substack{n \in \{1,\ldots,|\Sigma|\} \\ n \in \{1,\ldots,|S|\}}} R^s_{sum(n,l)}.$$

19: Find the corresponding system parameters: $\check{\mathbf{H}}_H = \check{\mathbf{H}}_{H(n,l)^*}$, $\check{\mathbf{H}}_L = \check{\mathbf{H}}_{L(n,l)^*}$, $\beta^*_L = \beta_{L(n,l)^*}$, $\beta^*_H = 1 - \beta^*_L$, $R^s_H\left(\check{A}\right) = R^s_{H(n,l)^*}$, $R^s_L\left(\check{A}\right) = R^s_{L(n,l)^*}$ and $R^s_{sum}\left(\check{A}\right) = R^s_{sum(n,l)^*}$.

Output: $\check{\mathbf{H}}_H, \check{\mathbf{H}}_L, \beta^*_H, \beta^*_L, R^s_H(\check{A}), R^s_L(\check{A})$, and $R^s_{sum}(\check{A})$.

Algorithm 2 UC-RAS-PA Scheme

Input: $K, M, M_s, \mathcal{P}, \mathcal{P}_{dif}, \sigma_n^2, \Omega, R_0, \alpha$, and $\mathbf{h}_k; k = 1, \ldots, K$.

1: Define $\Psi = [1, 2, \ldots, K]$ as the set of all active users, sorted according to their path loss parameters in ascending order, i.e. $L(d_1) < L(d_2) < \ldots < L(d_K)$.
2: Use the first T elements in Ψ to construct the set $\Phi = [1, \ldots, T]$ of HPC users and the last U elements to form the set $\Theta = [1, \ldots, U]$ of LPC users.
3: Construct the channel matrix associated with user clustering as:

$$\mathbf{H} = [\mathbf{H}_H \mathbf{H}_L] = [\mathbf{h}_1 \cdots \mathbf{h}_T \mathbf{h}_{T+1} \cdots \mathbf{h}_K] = [\mathbf{b}_1, \cdots, \mathbf{b}_M]^{\mathrm{T}}.$$

4: Define the set of all receive antennas as $\Upsilon = [1, \ldots, M]$ with $l \in \Upsilon$ indicating l^{th} antenna.
5: Calculate the power of l^{th} row $\mathbf{b}_l$ in $\mathbf{H}$ corresponding to l^{th} receive antenna as $\|\mathbf{b}_l\|^2$.
6: Sort the elements of Υ according to their associated powers in descending order.
7: Choose the first M_s elements in Υ which represent the best subset of receive antennas with highest powers.
8: Construct the subchannels of HPC and LPC associated with user clustering and RAS as $\check{\mathbf{H}}_H$ and $\check{\mathbf{H}}_L$, respectively.
9: Set $\mu = \left(\mathcal{P} - \mathcal{P}_{dif}\right)/2\mathcal{P}\Omega$, $\beta_L = 0$, and $R^s_{sum} = 0$.
10: **for** $q = 1$ *to* Ω **do**
11: Update $\beta_L = \beta_L + \mu$.
12: Calculate: $\beta_H = 1 - \beta_L$ and both terms R^s_H and R^s_L in the optimization problem (38).
13: **if** $\left(R^s_H \geq TR_0\right)$ AND $\left(R^s_L \geq UR_0\right)$ AND $\left(R^s_H + R^s_L > R^s_{sum}\right)$**then**
14: Update $R^s_{sum} = R^s_H + R^s_L$, $R^s_H\left(\check{A}\right) = R^s_H$, $R^s_L\left(\check{A}\right) = R^s_L$, and $\beta^*_L = \beta_L$.
15: **end if**
16: **end for**
17: Find the related system parameters as: $\beta^*_H = 1 - \beta^*_L$ and $R^s_{sum}\left(\check{A}\right) = R^s_H\left(\check{A}\right) + R^s_L\left(\check{A}\right)$.

Output: $\check{\mathbf{H}}_H, \check{\mathbf{H}}_L, \beta^*_H, \beta^*_L, R^s_H(\check{A}), R^s_L(\check{A})$, and $R^s_{sum}(\check{A})$.

A. RESULTS OF MODERAT SCALE MIMO-NOMA SCHEMES

In this part and without loss of generality, we consider $8 \times 12/4$, $8 \times 4/4$, $8 \times 12/6$, and $6 \times 12/4$ MIMO-NOMA configurations compared with the reference $4 \times 12/4$, $4 \times 4/4$, $6 \times 12/6$, 4, 4, 12/4, 4, 4, 4/4, 6, 2, 12/6, and 4, 2, 12/4 schemes.

In Fig. 2, capacity results of $8 \times 12/4$ scheme ($T = 4$; $U = 4$) using Algorithms 1-4 are shown as a function of SNR. This configuration demonstrate maximum connected users based on the utilized number of RFCs ($K = 2M_s = 8$), and RAS of M_s from $M = 3M_s$ spatial DoFs. Fig. 2(a) demonstrates the achieved overall sum rate $\left(R^s_{sum}\right)$ compared with the reference $4 \times 12/4$ and 4, 4, 12/4 schemes. As can be seen, performance of Algorithm 4 is very close to the optimal (Algorithms 1) and slightly outperforms Algorithms 3 and 2 for the entire range of SNRs. Besides and as expected, performance of all proposed algorithms is higher than that of $4 \times 12/4$ scheme due to $U = 4$ additional users and that of 4, 4, 12/4 scheme owing to clustering gain. Fig. 2(b) shows the average user rate in HPC of $T = 4$ and LPC of $U = 4$ as a function of SNR. It can be seen and for all designed algorithms that the performance of LPC users is

Algorithm 3 UC-PA-PBRAS Scheme

Input: $K, M, M_s, \mathcal{P}, \mathcal{P}_{dif}, \sigma_n^2, \Omega, R_0, \alpha$, and $\mathbf{h}_k; k = 1, \ldots, K$.

1: Define $\Psi = [1, 2, \ldots, K]$ as the set of all active users, sorted according to their path loss parameters in ascending order, i.e. $L(d_1) < L(d_2) < \ldots < L(d_K)$.
2: Use the first T elements in Ψ to construct the set $\Phi = [1, \ldots, T]$ of HPC users and the last U elements to form the set $\Theta = [1, \ldots, U]$ of LPC users.
3: Construct the channel matrix associated with user clustering as:

$$\mathbf{H} = [\mathbf{H}_H \mathbf{H}_L] = [\mathbf{h}_1 \cdots \mathbf{h}_T \mathbf{h}_{T+1} \cdots \mathbf{h}_K] = [\mathbf{b}_1, \cdots, \mathbf{b}_M]^{\mathrm{T}}.$$

4: Set $\mu = \left(\mathcal{P} - \mathcal{P}_{dif}\right) / 2\mathcal{P}\Omega$, $\beta_L = 0$, and $R_{sum} = 0$.
5: **for** $q = 1$ *to* Ω **do**
6: Update $\beta_L = \beta_L + \mu$.
7: Calculate: $\beta_H = 1 - \beta_L$, R_H using (22), and R_L using (23).
8: **if** $(R_H \geq TR_0)$ AND $(R_L \geq UR_0)$ AND $(R_H + R_L > R_{sum})$ **then**
9: Update $R_{sum} = R_H + R_L$ and $\beta_L^* = \beta_L$.
10: **end if**
11: **end for**
12: Calculate: $\beta_H^* = 1 - \beta_L^*$.
13: Define the set of all receive antennas as $\Upsilon = [1, \ldots, M]$ with $l \in \Upsilon$ indicating l^{th} antenna.
14: Calculate the power of l^{th} row $\mathbf{b}_l$ in $\mathbf{H}$ corresponding to l^{th} receive antenna as $\|\mathbf{b}_l\|^2$.
15: Sort the elements of Υ according to their associated powers in descending order and choose the first M_s elements as the best subset of receive antennas of highest powers.
16: Construct the subchannels of HPC and LPC associated with user clustering and RAS as $\check{\mathbf{H}}_H$ and $\check{\mathbf{H}}_L$, respectively.
17: Find the corresponding system parameters as: $R_H^s\left(\check{A}\right)$ using (29), $R_L^s\left(\check{A}\right)$ using (30), and $R_{sum}^s\left(\check{A}\right) = R_H^s\left(\check{A}\right) + R_L^s\left(\check{A}\right)$.

Output: $\check{\mathbf{H}}_H, \check{\mathbf{H}}_L, \beta_H^*, \beta_L^*, R_H^s(\check{A}), R_L^s(\check{A})$, and $R_{sum}^s(\check{A})$.

Algorithm 4 UC-PA-CBRAS Scheme

Input: $K, M, M_s, \mathcal{P}, \mathcal{P}_{dif}, \sigma_n^2, \Omega, R_0, \alpha$, and $\mathbf{h}_k; k = 1, \ldots, K$.

1: Define $\Psi = [1, 2, \ldots, K]$ as the set of all active users, sorted according to their path loss parameters in ascending order, i.e. $L(d_1) < L(d_2) < \ldots < L(d_K)$.
2: Use the first T elements in Ψ to construct the set $\Phi = [1, \ldots, T]$ of HPC users and the last U elements to form the set $\Theta = [1, \ldots, U]$ of LPC users.
3: Construct the channel matrix associated with user clustering as:

$$\mathbf{H} = [\mathbf{H}_H \mathbf{H}_L] = [\mathbf{h}_1 \cdots \mathbf{h}_T \mathbf{h}_{T+1} \cdots \mathbf{h}_K] = [\mathbf{b}_1, \cdots, \mathbf{b}_M]^{\mathrm{T}}.$$

4: Set $\mu = \left(\mathcal{P} - \mathcal{P}_{dif}\right) / 2\mathcal{P}\Omega$, $\beta_L = 0$, and $R_{sum} = 0$.
5: **for** $q = 1$ *to* Ω **do**
6: Update $\beta_L = \beta_L + \mu$.
7: Calculate: $\beta_H = 1 - \beta_L$, R_H using (22), and R_L using (23).
8: **if** $(R_H \geq TR_0)$ AND $(R_L \geq UR_0)$ AND $(R_H + R_L > R_{sum})$ **then**
9: Update $R_{sum} = R_H + R_L$ and $\beta_L^* = \beta_L$.
10: **end if**
11: **end for**
12: Calculate: $\beta_H^* = 1 - \beta_L^*$.
13: Define the set of all receive antennas as $\Upsilon = [1, \ldots, M]$ with $l \in \Upsilon$ indicating l^{th} antenna.
14: Use Υ to construct the set of all potential receive antenna subsets $S = \left\{\mathcal{S}_1, \ldots, \mathcal{S}_{|S|}\right\}$ with $|S| = M! / (M_s!\,(M - M_s)!)$.
15: **for** $l = 1$ *to* $|S|$ **do**
16: Construct both $\check{\mathbf{H}}_{H(l)}$ and $\check{\mathbf{H}}_{L(l)}$ based on $\mathcal{S}_l$ to find $R_{H(l)}^s$ using (29), $R_{L(l)}^s$ using (30), and $R_{sum(l)}^s = R_{H(l)}^s + R_{L(l)}^s$.
17: **end for**
18: Choose the index that maximize the sum rate associated with user clustering and RAS as: $l^* = \arg\max_{l \in \{1,\ldots,|S|\}} R_{sum(l)}^s$.
19: Find the corresponding system parameters as: $\check{\mathbf{H}}_H = \check{\mathbf{H}}_{H(l^*)}$, $\check{\mathbf{H}}_L = \check{\mathbf{H}}_{L(l^*)}$, $R_H^s\left(\check{A}\right) = R_{H(l^*)}^s$, $R_L^s\left(\check{A}\right) = R_{L(l^*)}^s$, and $R_{sum}^s\left(\check{A}\right) = R_{sum(l^*)}^s$.

Output: $\check{\mathbf{H}}_H, \check{\mathbf{H}}_L, \beta_H^*, \beta_L^*, R_H^s(\check{A}), R_L^s(\check{A})$, and $R_{sum}^s(\check{A})$.

increased as the SNR increases while results of HPC users outperforms those of LPC at low to moderate SNRs. It then demonstrate saturation at high SNRs due to higher LPC interference compared with the receiver noise power. In this case, the intersections between achieved HPC and LPC curves demonstrate the equal rate points of connected users (i.e. user-fairness of equal rates) based on the utilized algorithms. For instance, each served user will attain 3.68 bit/s/Hz using the optimal solution (Algorithm 1) at SNR of 27dB whereas sub-optimal methods (Algorithms 2-4) show very close and less results. Other unequal user rate points can be selected based on the per-user QoS requirements for given SNR targets.

For $8 \times 4/4$ scheme of $K = 2M_s = 8$ users without RAS ($M = M_s = 4$), the capacity outcomes of proposed algorithms are shown in Fig.3 as a function of SNR. In Fig. 3(a), the sum rates are presented compared with the reference $4 \times 4/4$ and 4, 4, 4/4 schemes. It can be seen clearly that the performance curves of Algorithms 2-4 are identical due to the absence of RAS and similar adopted mechanisms of user clustering and power allocation in these methods. But,

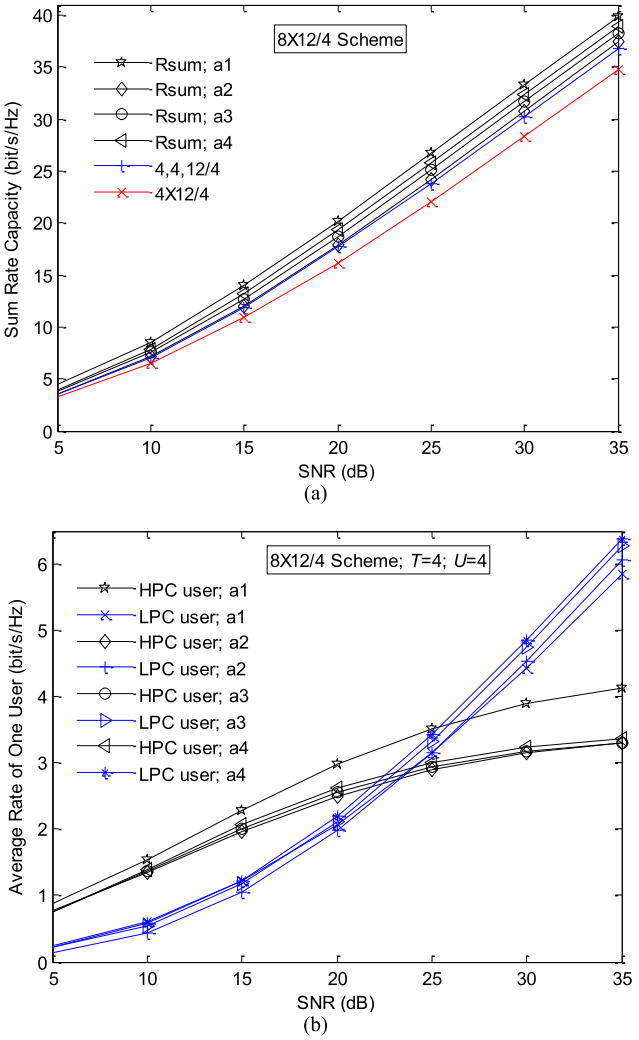

FIGURE 2. The capacity of 8 × 12/4 scheme using Algorithms 1-4 in bit/s/Hz as a function of SNR: (a) Achieved sum rate compared with the reference 4 × 12/4 and 4, 4, 12/4 schemes. (b) Average user rate in HPC and LPC.

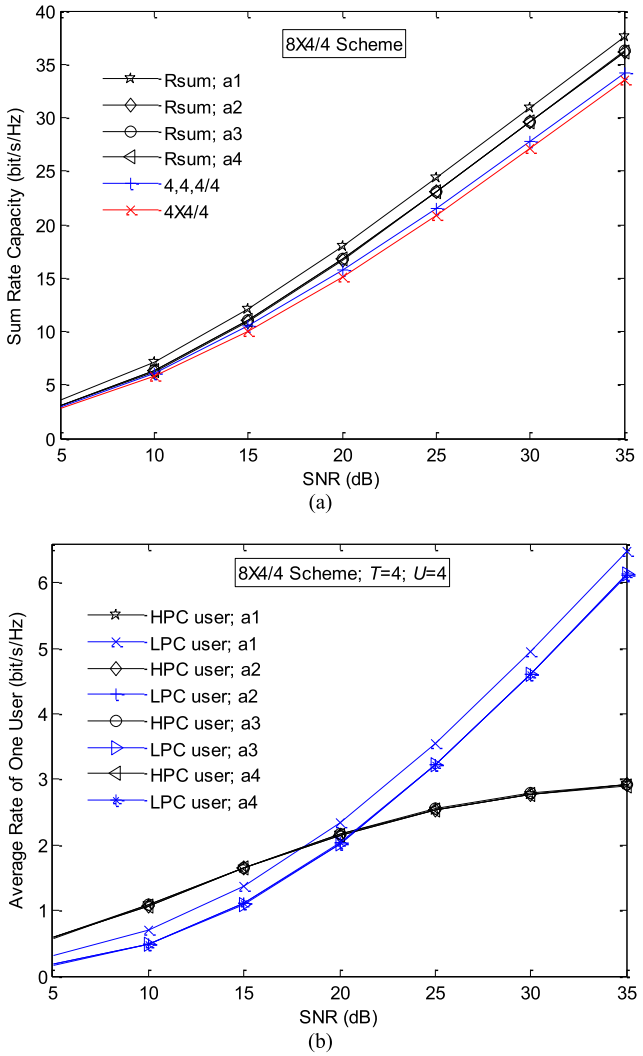

FIGURE 3. The capacity of 8 × 4/4 scheme using algorithms 1-4 in bit/s/Hz as a function of SNR. (a) Achieved sum rate compared with the reference 4 × 4/4 and 4, 4, 4/4 schemes. (b) Average user rate in HPC and LPC.

Algorithm 1 still provides the optimal R^s_{sum} through joint user clustering and power allocation process. For example, it achieves 24.34 bit/s/Hz at SNR of 25 dB higher than the other methods by 1.34 bit/s/Hz and the benchmark 4, 4, 4/4 by 2.86 bit/s/Hz. Accordingly, the performance gap between Algorithm 1 and other algorithms is reflected to R^s_L rather than R^s_H since the decoding process of LPC signals (after SIC of HPC signals) is interference-free in contrast to that of HPC which suffers from the interference of LPC signals. This is shown in Fig. 3(b) of the achieved average user rates for HPC and LPC users. In this case, Algorithms 2-4 demonstrate equal user rate point of 2.23 bit/s/Hz at SNR = 20.8dB while Algorithm 1 attains 1.96 bit/s/Hz at less SNR of 18.1 dB.

In Fig. 4, capacity results of 8 × 12/6 scheme ($T = 6$; $U = 2$) using more RFCs ($M_s = 6$) are shown as a function of SNR. This configuration demonstrates the case of $K = 8$, less than maximum allowed number of $2M_s = 12$ users (i.e. low user overloading of $U = 2$ in LPC), and RAS of M_s from $M = 2M_s$ spatial DoFs. Fig. 4(a) shows the achieved sum rates compared with the reference 6 × 12/6 and 6, 2, 12/6 schemes. As can be seen, the achieved results of Algorithms 2 and 3 are the same for the entire range of SNRs due to less utilized spatial DoFs. In addition, results of Algorithm 4 outperform that of Algorithms 2 and 3, and become more closer to the optimal solution (Algorithms 1). Performance of all algorithms is also higher than the reference 6 × 12/6 scheme owing to $U = 2$ additional users and that of 6, 2, 12/6 scheme due to user clustering gain. Fig. 4(b) illustrates the average user rate in HPC and LPC as a function of SNR. As can be seen and for all utilized algorithms, the user rate in both clusters grows as the SNR increases, and LPC users perform better than HPC at moderate to high SNRs since they benefit from interference-free decoding. Besides, the curves of HPC users are not saturated at high SNRs due to low interference level from LPC compared with

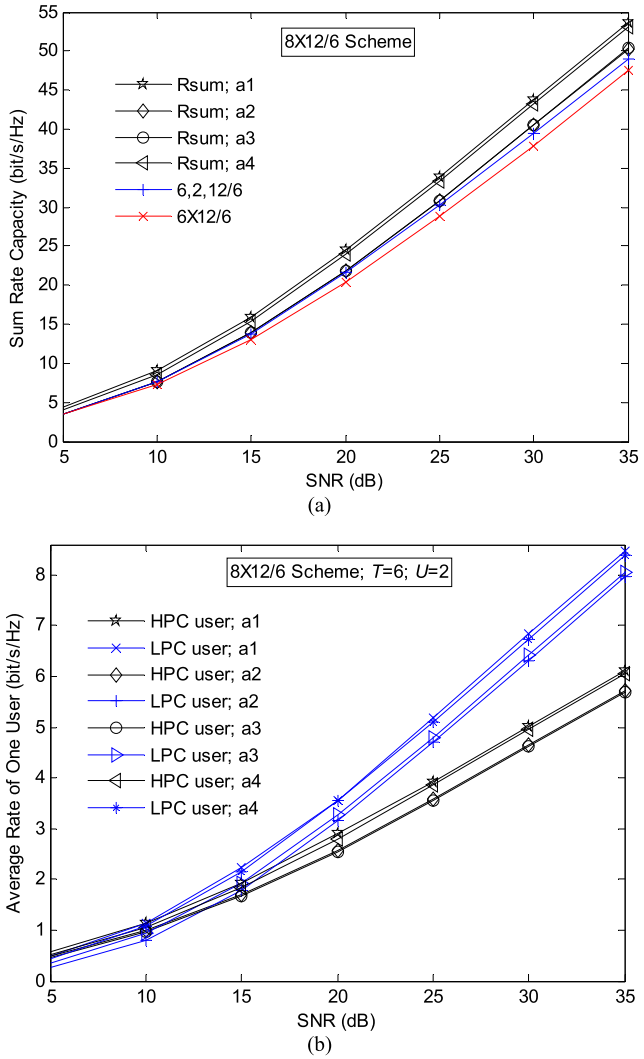

FIGURE 4. The capacity of 8 × 12/6 scheme using algorithms 1-4 in bit/s/Hz as a function of SNR. (a) Achieved sum rate compared with the reference 6 × 12/6 and 6, 2, 12/6 schemes. (b) Average user rate in HPC and LPC.

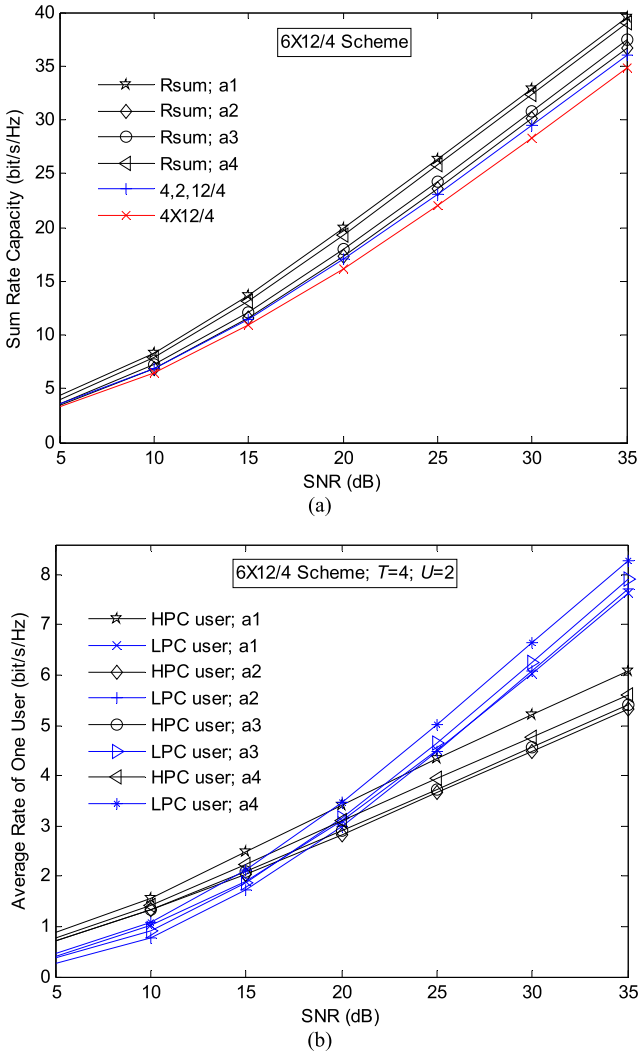

FIGURE 5. The capacity of 6 × 12/4 scheme using algorithms 1-4 in bit/s/Hz as a function of SNR. (a) Achieved sum rate compared with the reference 4 × 12/4 and 4, 2, 12/4 schemes. (b) Average user rate in HPC and LPC.

the noise power. In this case, close equal user rate points for Algorithms 1-4 are achieved.

Capacity results of $6 \times 12/4$ scheme are presented in Fig. 5 as a function of SNR compared with the reference $4 \times 12/4$ and 4, 2, 12/4 schemes. This configuration serves $K = 6$ users ($T = 4$; $U = 2$) less than the allowed number ($2M_s = 8$) and employs $M = 3M_s$. Summary of the achieved capacity results (bit/s/Hz) of considered schemes in Figs. 2(a), 3(a), 4(a), and 5(a) using proposed algorithms at SNR of 25 dB is shown in Table 3 compared with the associated reference systems. Furthermore, summary of equal user rate results (bit/s/Hz) and associated SNRs (dB) for considered schemes in Figs. 2(b), 3(b), 4(b), and 5(b) are presented in Table 4.

Capacity regions of above scenarios are also shown in Figs. 6 and 7 for SNR of 20, 25, and 30 dB. The carried out results for $8 \times M/4$ schemes of full user overloading ($T = 4$; $U = 4$) are presented in Fig. 6(a) using $M = 3M_s$ and Fig. 6(b) for the case of $M = M_s$. As can be seen for $M = 12$, the maximum clusters' sum rates at the operating point $\breve{A} = \left(R_H^s, R_L^s\right)$ (please refer to Fig. 1) and hence the associated user rate and overall sum rate $\left(R_{sum}^s = R_H^s + R_L^s\right)$ depend on the utilized algorithm and SNR. The optimal R_{sum}^s performance (line between corner points $\breve{A}\breve{B}$) is achieved using Algorithm 1 while Algorithm 2 of lowest complexity demonstrate somewhat less performance than the others over considered SNRs. For instance at SNR = 20dB, the results (point $\breve{A}$) of Algorithms 1 and 2 are $\left(R_H^s = 11.9, R_L^s = 8.28\right)$ and $\left(R_H^s = 10, R_L^s = 7.9\right)$, respectively. Thus, HPC users have the benefit of more rate (QoS) compared with LPC users. On the other hand at SNR = 30dB, the achieved points using Algorithms 1 and 2 are $\left(R_H^s = 15.58, R_L^s = 17.69\right)$ and $\left(R_H^s = 12.61, R_L^s = 18.18\right)$, respectively. Thus, LPC users enjoys more rate than HPC users. Summary of the realized outcomes is

TABLE 3. Summary of capacity results of considered schemes in Figs. 2(a), 3(a), 4(a), and 5(a) using proposed algorithms at SNR of 25 dB. The sum rates of 6, 2, 12/6, 4, 4, 12/4, 4, 2, 12/4, and 4, 4, 4/4 reference schemes are 30.25, 23.80, 23.10, and 21.52 bit/s/Hz, respectively.

Scheme	Performance Measure	Capacity of Proposed Algorithms in bit/s/Hz at SNR of 25 dB			
		Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4
$8 \times 12/4$ $(T = 4; U = 4)$	R^s_{sum}	26.68	24.25	25.09	25.76
	R^s_H	14.08	11.61	11.75	12.05
	R^s_L	12.60	12.64	13.34	13.71
$8 \times 4/4$ $(T = 4; U = 4)$	R^s_{sum}	24.38	23.04	23.04	23.04
	R^s_H	14.21	12.91	12.91	12.91
	R^s_L	10.17	10.13	10.13	10.13
$8 \times 12/6$ $(T = 6; U = 2)$	R^s_{sum}	33.96	30.90	30.90	33.38
	R^s_H	23.58	21.46	21.31	23.16
	R^s_L	10.38	9.44	9.59	10.22
$6 \times 12/4$ $(T = 4; U = 2)$	R^s_{sum}	26.35	23.62	24.27	25.79
	R^s_H	17.34	14.64	14.95	15.76
	R^s_L	9.01	8.98	9.32	10.03

TABLE 4. Summary of equal user rates and associated SNRs (dB) of considered schemes in Figs. 2(b), 3(b), 4(b), and 5(b) using proposed algorithms.

Scheme	Equal User Rate Point (bit/s/Hz) and Associated SNR (dB)							
	Algorithm 1		Algorithm 2		Algorithm 3		Algorithm 4	
	Rate	SNR	Rate	SNR	Rate	SNR	Rate	SNR
$8 \times 12/4; (T = 4; U = 4)$	3.68	27.0	2.76	23.4	2.75	22.6	2.82	22.5
$8 \times 4/4; (T = 4; U = 4)$	1.96	18.1	2.23	20.8	2.23	20.8	2.23	20.8
$8 \times 12/6; (T = 6; U = 2)$	1.15	9.3	1.50	13.5	1.03	10.3	0.96	9.2
$6 \times 12/4; (T = 4; U = 2)$	4.00	23.2	2.57	18.4	2.50	17.5	2.45	16.2

TABLE 5. Summary of capacity results of considered schemes in Figs. 6(a) and 7(a) using proposed algorithms at SNR of 20 dB and 30 dB.

Scheme	SNR (dB)	Performance Measure	Capacity of Proposed Algorithms in bit/s/Hz			
			Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4
$8 \times 12/4$ $(T = 4; U = 4)$	20	R^s_{sum}	20.18	17.90	18.69	19.30
		R^s_H	11.90	10.00	10.22	10.51
		R^s_L	8.28	7.90	8.47	8.79
	30	R^s_{sum}	33.27	30.79	31.64	32.35
		R^s_H	15.58	12.61	12.68	12.97
		R^s_L	17.69	18.18	18.96	19.38
$8 \times 12/6$ $(T = 6; U = 2)$	20	R^s_{sum}	24.51	21.82	21.85	23.94
		R^s_H	17.38	15.51	15.33	16.85
		R^s_L	7.13	6.31	6.52	7.09
	30	R^s_{sum}	43.75	40.46	40.50	43.16
		R^s_H	30.08	27.82	27.68	29.69
		R^s_L	13.67	12.64	12.82	13.47

presented in Table 5. Notice that similar conclusions can be found from Fig. 6(b) except that Algorithms 2-4 have the same performance due to absence of RAS diversity and similar user clustering and power allocation. The equal rate points and associated SNRs of these schemes are shown in Table 4.

Capacity regions of $K \times 12/M_s$ schemes with partial user overloading of $U = 2$ are shown in Figs. 7(a) and 7(b) using $8 \times 12/6$ with $(M = 2M_s)$ and $6 \times 12/4$ with $(M = 3M_s)$, respectively. It can be seen that R^s_H performance of all algorithms is better than R^s_L due to low interference level from LPC users. Nevertheless, LPC users attain more rate than HPC users due to low user overloading, except the optimal region of $6 \times 12/4$ using Algorithm 2 at SNR of 20dB (see Fig. 5(b)). Summary of the achieved results of $8 \times 12/6$ scheme is given in Table 5 while the equal rate points and associated SNRs of these schemes are shown earlier in Table 4.

B. RESULTS OF LARGE SCALE MIMO-NOMA SCHEMES

In this part, we consider $K \times 160/M_s$ schemes with fixed number of $M = 160$ antennas to demonstrate the impact

TABLE 6. Summary of capacity results of considered 80 × 160/M_s schemes in Figs. 8(a) and 9(a) using Algorithm 2 at SNR of 30 dB.

Performance Measure	Capacity of $80 \times 160/M_s$ schemes in bit/s/Hz for different number of RFCs (M_s)			
	$80 \times 160/40$ $(T = 40; U = 40)$	$80 \times 160/50$ $(T = 50; U = 30)$	$80 \times 160/60$ $(T = 60; U = 20)$	$80 \times 160/70$ $(T = 70; U = 10)$
R^s_{sum}	169.8	195.2	214.3	228.8
R^s_H	88.2	117.7	149.4	192.4
R^s_L	81.6	77.5	64.9	36.4

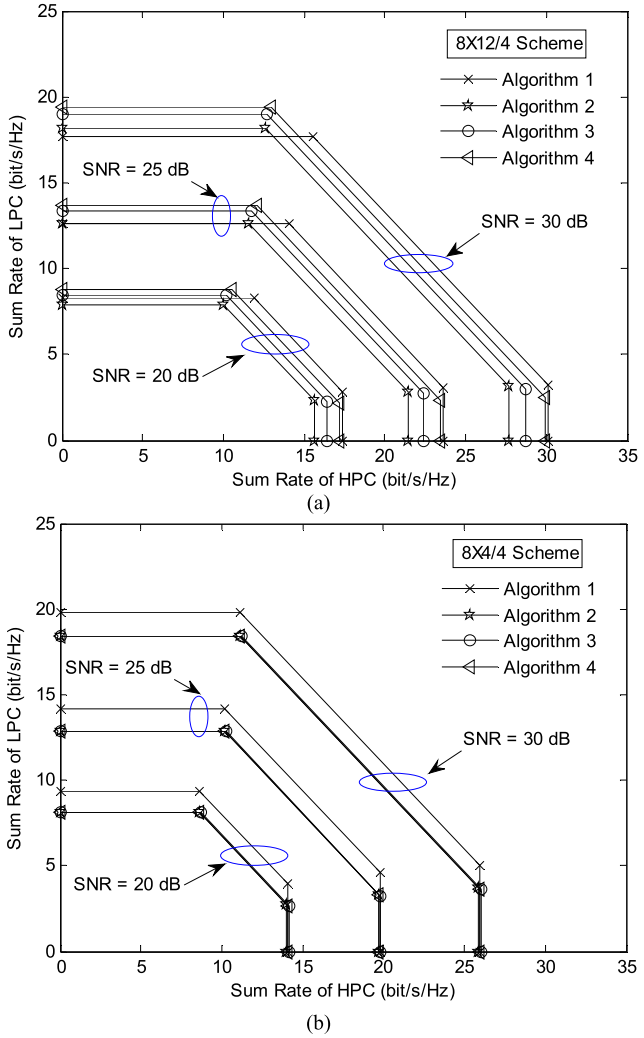

FIGURE 6. Capacity region of 8 × $M/4$ schemes using algorithms 1-4 at SNR = 20, 25, and 30dB: (a) $M = 3M_s = 12$. (b) $M = M_s = 4$.

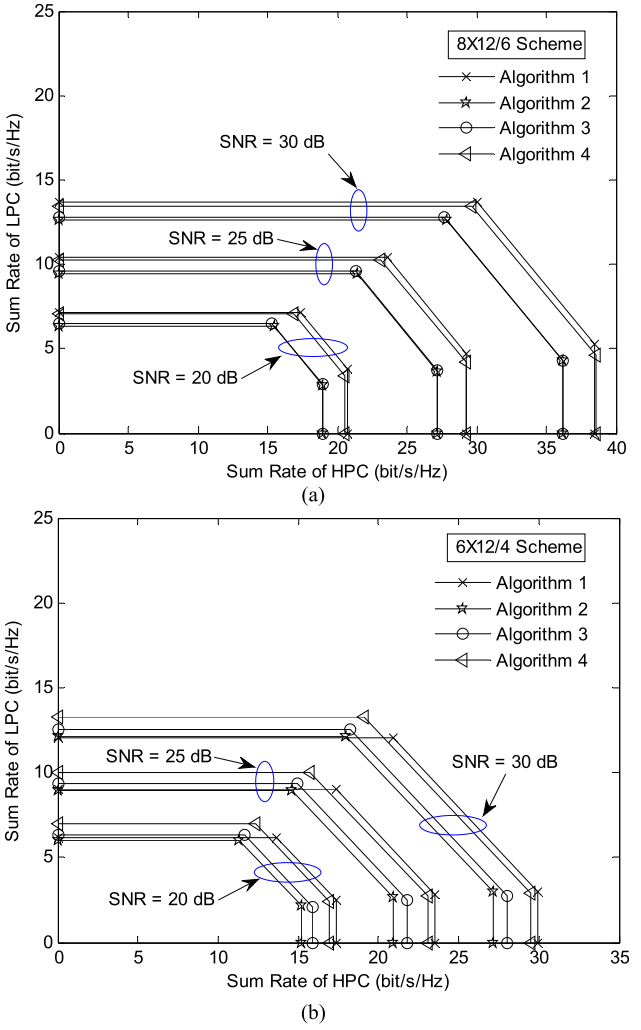

FIGURE 7. Capacity region of K × 12/M_s schemes using algorithms 1-4 at SNR = 20, 25, and 30dB: (a) $K = 8$ and $M_s = 6$. (b) $K = 6$ and $M_s = 4$.

of different user connectivity (K) and implemented number of RFCs (M_s) (consequently, diverse user overloading (U) and/or M/M_s ratios). In particular, two important scenarios are investigated as $80 \times 160/M_s$ and $K \times 160/60$. For all implemented schemes, we found that the results of both low and moderate complexity algorithms (2 and 3, respectively) are very tight to each other whereas the computational efforts of Algorithms 1 and 4 are extremely high and cannot be executed practically. Therefore, we present the results of lowest complexity technique (Algorithm 2) in Figs. 8–11.

In Fig. 8(a), the sum rates of $80 \times 160/M_s$ schemes are presented as a function of SNR compared with the benchmark systems $40, 40, 160/70$, $40, 40, 160/40$, $70 \times 160/70$, and $40 \times 160/40$. This set-up demonstrates a fixed number of connected users ($K = 80$) and different M/M_s ratios based on the number of RFCs ($M_s = 70$ and 40). As can be seen, the sum rate is increased considerably as M_s increases (thus, U decreases) for moderate to high SNRs. For example at SNR = 30dB, the achieved results are 169.8 and 228.8 bit/s/Hz for $M_s = 40$ and $M_s = 70$, respectively. Besides, these sum rates are higher than the associated references owing to the additional users and diversity of user clustering. Fig. 8(b) shows the sum rate of $K \times 160/60$ schemes as a function of SNR com-

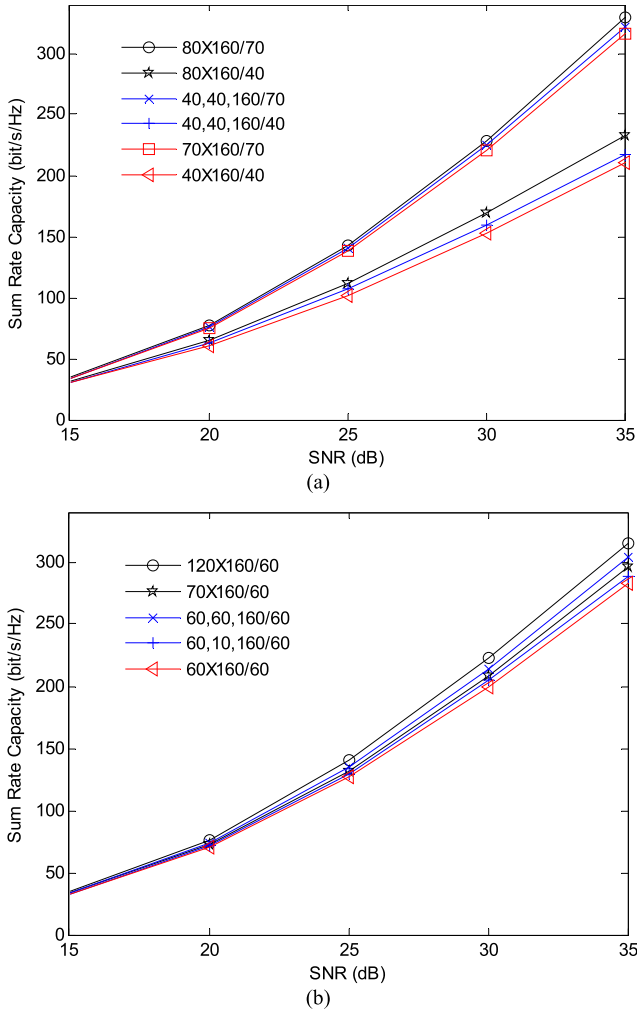

FIGURE 8. Sum rate capacity of $K \times 160/M_s$ schemes using Algorithm 2 in bit/s/Hz as a function of SNR. (a) Results for $K = 80$ and $M_s = 70$ and 40 compared with the references 40, 40, 160/70, 40, 40, 160/40, $70 \times 160/70$, and $40 \times 160/40$; (b) Results for $M_s = 60$ with $K = 120$ and 70 compared with the references 60, 60, 160/60, 60, 10, 160/60, and $60 \times 160/60$.

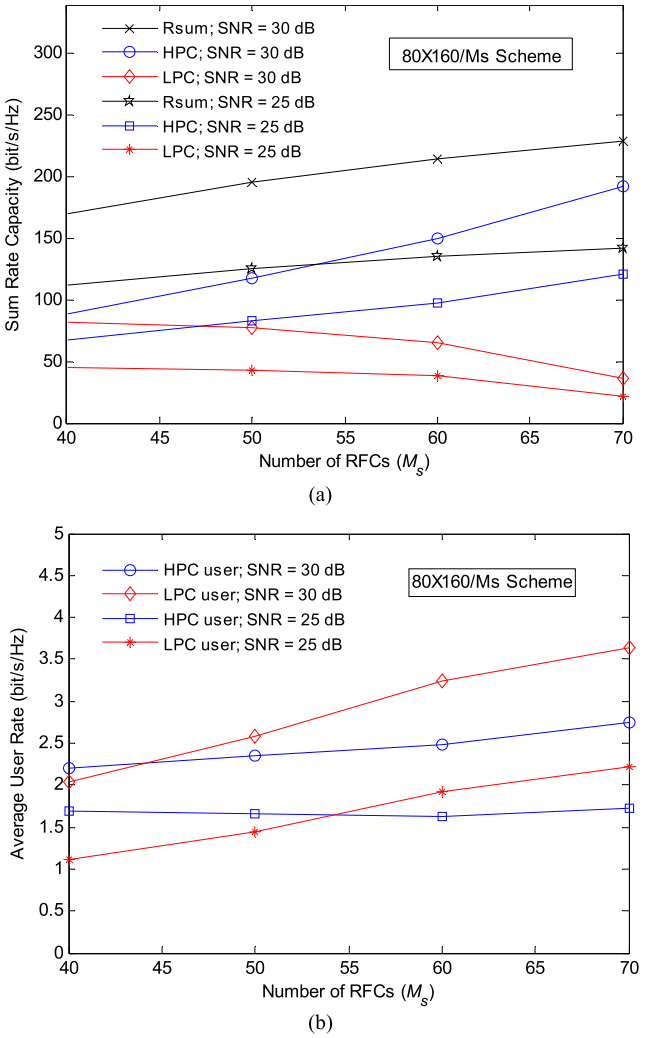

FIGURE 9. The capacity of $80 \times 160/M_s$ schemes using algorithm 2 in bit/s/Hz as a function of the number of RFCs (M_s) and for SNR = 25 and 30dB. (a) Sum rate capacity of the considered schemes, HPC, and LPC. (b) Average user rate in HPC and LPC.

pared with the references 60, 60, 160/60, 60, 10, 160/60, and $60 \times 160/60$. This scenario demonstrates a fixed number of $M_s = 60$ RFCs and different number of users as $K = 120$ and 70. It can be seen that the performance is increased considerably as K increases (i.e. U increases) for moderate to high SNRs and higher than the associated reference systems. Summary of the achieved results are presented in Tables 6 and 7.

In Fig. 9(a), the sum rates R^s_{sum}, R^s_H, and R^s_L of $80 \times 160/M_s$ schemes are shown as a function of M_s for SNR of 25 and 30 dB. It is clear that as M_s increased, the number of HPC users ($T = M_s$) will be increased accordingly whereas the number of LPC users (U) will be decreased and hence producing less interference to HPC users. This allows R^s_H to grow considerably in contrast to R^s_L which depends on the allocated users in LPC despite the interference-free decoding. For instance when $M_s = 40$ and 70, the achieved results of R^s_H at SNR = 30 dB are 88.2 and 192.4 bit/s/Hz, respectively while 81.6 and 36.4 bit/s/Hz are shown for R^s_L, respectively. Notice that as M_s increased, the average user rate in LPC is increased more than HPC as illustrated in Fig. 9(b). Furthermore, the equal user rate point depends on the target SNR. For example, equal user rate of 1.64 bit/s/Hz is achieved at SNR = 25 dB using $M_s = 54$ RFCs ($U = 26$) compared to 2.27 bit/s/Hz at SNR = 30 dB with less number of $M_s = 44$ RFCs ($U = 36$). This set-up reveal the valuable trade-offs between achieved overall sum rate, user-fairness (equal user rate), system complexity (M_s), user overloading (U), and SNR. Summary of the achieved results are presented in Tables 6 and 8.

In Fig. 10(a), the sum rates R^s_{sum}, R^s_H, and R^s_L of $K \times 160/60$ configurations are presented as a function of K for SNR of 25 and 30 dB. In this scenario as K increased, the number of LPC users increases and hence producing more interference to HPC users. Accordingly, R^s_H drops noticeably in contrast to R^s_L that take advantage of more users (U) and interference-free decoding. It can be seen also from Fig. 10(b) that as K increased, the average user rate in both HPC and LPC is

TABLE 7. Summary of capacity results of considered K × 160/60 schemes in Figs. 8(b) and 10(a) using algorithm 2 at SNR of 30 dB.

Performance Measure	Capacity of $K \times 160/60$ schemes in bit/s/Hz for different number of users (K)				
	$61 \times 160/60$ $(T = 60; U = 1)$	$70 \times 160/60$ $(T = 60; U = 10)$	$90 \times 160/60$ $(T = 60; U = 30)$	$110 \times 160/60$ $(T = 60; U = 50)$	$120 \times 160/60$ $(T = 60; U = 60)$
R^s_{sum}	200.9	208.3	217.8	221.6	222.3
R^s_H	196.6	170.2	138.4	125.7	121.6
R^s_L	4.3	38.1	79.4	95.9	100.7

TABLE 8. Summary of equal user rates (bit/s/Hz) of considered schemes in Figs. 9(b) and 10(b) using Algorithm 2 at SNR of 25 dB and 30 dB.

Scheme	Equal User Rate Point (bit/s/Hz)	
	SNR = 25 dB	SNR = 30 dB
$80 \times 160/M_s;\ (T = M_s; U = 80 - T)$	1.64 (using $M_s = 54$ RFCs)	2.27 (using $M_s = 44$ RFCs)
$K \times 160/60;\ (T = 60; U = K - 60)$	1.58 (for $K = 88$ users)	2.16 (for $K = 102$ users)

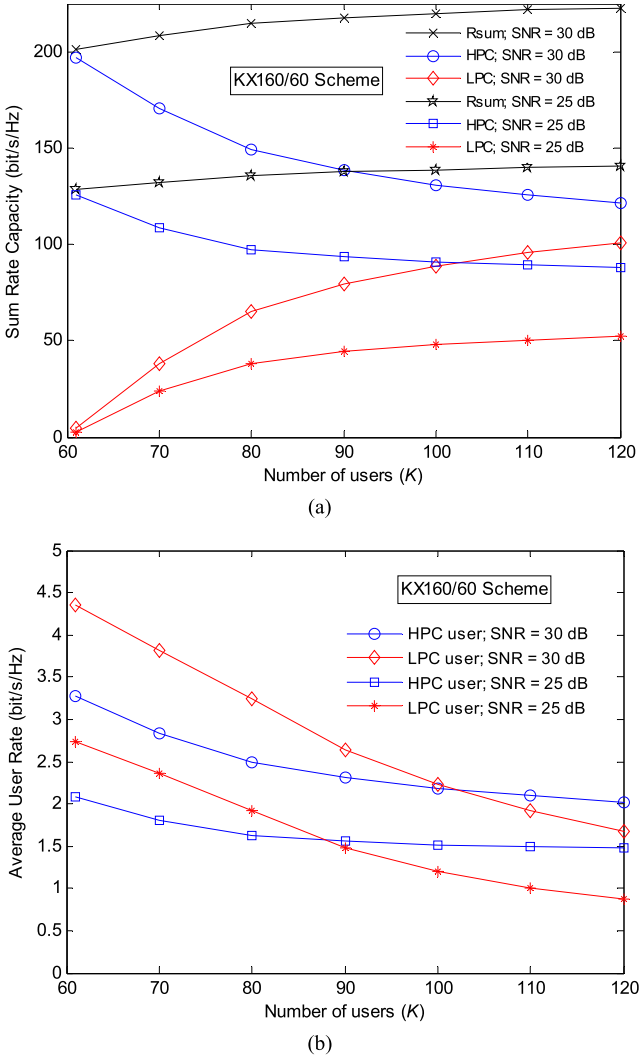

FIGURE 10. The capacity of $K \times 160/60$ schemes using Algorithm 2 in bit/s/Hz as a function of K and for SNR = 25 and 30dB. (a) Sum rate capacity of considered schemes, HPC, and LPC. (b) Average user rate in HPC and LPC.

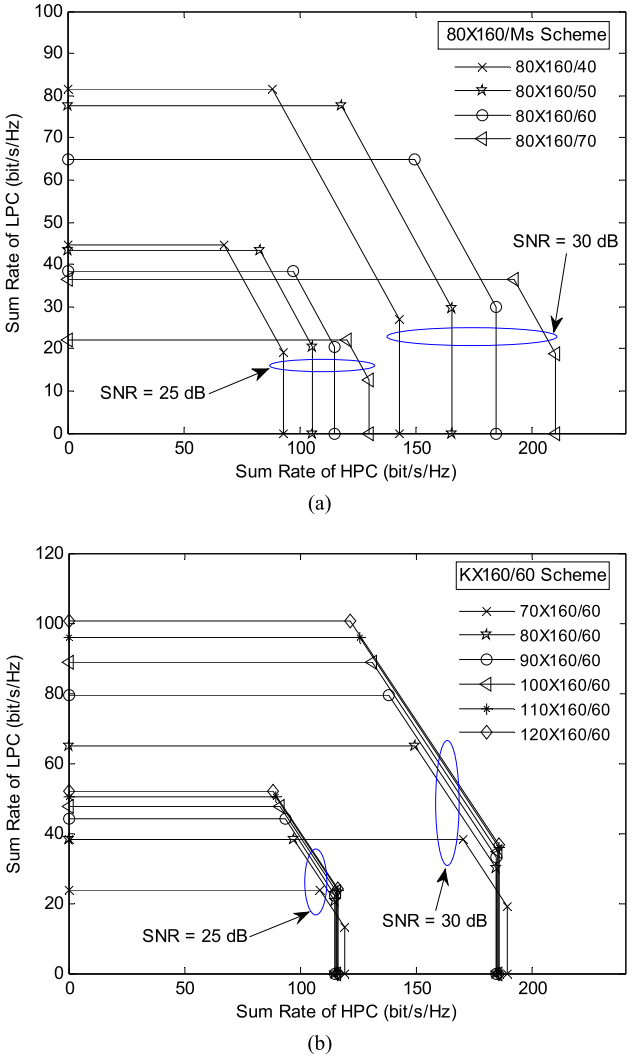

FIGURE 11. Capacity region of $K \times 160/M_s$ schemes using algorithm 2 in bit/s/Hz at SNR = 25 and 30dB: (a) $K = 80$ and $M_s = 40$, 50, 60, and 70. (b) $M_s = 60$ and $K = 70$, 80, 90, 100, 110, and 120.

decreased, and the equal user rate point depends also on the utilized SNR. For instance, equal user rate of 1.58 bit/s/Hz is realized at SNR = 25 dB using $K = 88$ ($U = 28$) compared to 2.16 bit/s/Hz at SNR = 30 dB with more number of $K = 102$ users ($U = 42$). This scenario also make obvious of the important tradeoffs between achieved overall sum rate, user-fairness, user overloading, and SNR. Summary of the achieved results are presented in Tables 7 and 8.

Finally, capacity regions of considered $80 \times 160/M_s$ and $K \times 160/60$ schemes are shown for SNR of 25 and 30 dB in Fig. 11(a) and Fig. 11(b), respectively. The presented results are inline with those in Fig. 8, 9(a), and 10(a), where for a given SNR, the maximum achievable clusters' sum rates at operating point $\check{A} = \left(R_H^s, R_L^s\right)$, and hence the associated user rate and overall sum rate, depend on the operating SNR, numbers of M_s RFCs, and supported users K (see Tables 6 and 7 for SNR $= 30$ dB). For instance and for both SNRs, HPC users have the benefit of more rates compared with LPC users when $M_s = 40$ and vice versa for $M_s = 70$. In Fig. 11(b) and based on the results of point $\check{A}$, we can see that LPC users enjoy higher rates than HPC users when $K = 70$ and conversely for $K = 120$ users.

The presented results validate that the proposed massive MIMO-NOMA scheme offers high user connectivity and overall sum rate, flexible clusters' sum rate distributions for given SNR and desired QoS, and significantly low system complexity. For instance, users of weak channel conditions can be served fairly as those of high channel conditions with equal rates. Thus, the inherent near-far problem in cellular systems can be efficiently exploited to allow more users, and hence less latency. Note that for the existing cellular systems, transmit power of weak users (e.g. near the cell edge) should be raised to enhance their data rate, but at cost of high co-channel interference with the other cells, which affects the overall network performance.

VI. CONCLUSION

This paper investigated massive MIMO-NOMA with RAS as a promising technology for 5G cellular networks to deliver higher connectivity and sum rate capacity, improved user-fairness, and least implementation complexity. In this scheme, simultaneous transmission of the MU-MIMO signals from HPC and LPC is attained using NOMA with efficient power allocation policy. For the performance evaluation, the sum rate and capacity region expressions have been derived for the uplink Rayleigh fading channel, and the optimal operating point that maximizes the overall sum rate is demonstrated. The optimal solution for the key problem of dynamic user clustering, RAS, and power allocation has been presented based on the exhaustive search for overall sum rate maximization using OJUC-RAS-PA (Algorithm 1) under received power constraints and minimum users' rates targets. Owing to unaffordable complexity of the optimal algorithm for large scale schemes, feasible algorithms, namely UC-RAS-PA (Algorithm 2), UC-PA-PBRAS (Algorithm 3), and UC-PA-CBRAS (Algorithm 4), have been proposed with sub-optimal performance by splitting the joint problem into low complexity components. Least computational efforts of $\mathcal{O}\left(K^2M + M_s^3\Omega\right)$ is achieved when Algorithm 2 is employed. Simulations results of different moderate and large scale MIMO-NOMA scenarios validated the effectiveness of the designed low complexity algorithms compared with the optimal solution and conventional schemes. It has been demonstrated that a significant increase in connected users, up to two-fold for the utilized RFCs ($\leq 2M_s$), can be achieved with higher overall sum rate capacity and desired user-fairness of equal/unequal rates. In addition, valuable tradeoffs can be realized by controlling the optimal operating point on the capacity region at target SNR using the main complexity- related parameters (i.e. U, M_s, and M). In future work, the impact of imperfect CSI and user synchronization on the performance of proposed scheme will be investigated.

REFERENCES

[1] Z. Ding *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[2] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[3] M. Patzold, "Countdown for the full-scale development of 5G new radio [mobile radio]," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 7–13, Jun. 2018.

[4] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Commun. Standards Mag.*, vol. 1, no. 4, pp. 24–30, Dec. 2017.

[5] H. Ji *et al.*, "Overview of full-dimension MIMO in LTE-advanced pro," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 176–184, Feb. 2017.

[6] L. Lei, D. Yuan, and P. Värbrand, "On power minimization for non-orthogonal multiple access (NOMA)," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2458–2461, Dec. 2016.

[7] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5G new radio: Waveform, frame structure, multiple access, and initial access," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, Jun. 2017.

[8] E. G. Larsson, T. L. Marzetta, H. Q. Ngo, and H. Yang, "Antenna count for massive MIMO: 1.9 GHz vs. 60 GHz," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 132–137, Sep. 2018.

[9] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[10] Y. Gao, H. Vinck, and T. Kaiser, "Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1346–1360, Mar. 2018.

[11] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.

[12] P. V. Amadori and C. Masouros, "Interference-driven antenna selection for massive multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 5944–5958, Aug. 2016.

[13] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Multi-switch for antenna selection in massive MIMO," in *Proc. IEEE GLOBCOM Conf.*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[14] W. A. Al-Hussaibi and F. H. Ali, "Group layer MU-MIMO for 5G wireless systems," *Telecommun. Syst.*, vol. 6, pp. 1–16, Jan. 2019. doi: 10.1007/s11235-018-00536.

[15] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[16] N. B. Mehta, S. Kashyap, and A. F. Molisch, "Antenna selection in LTE: From motivation to specification," *IEEE Commun. Mag.*, vol. 50, no. 10, pp. 144–150, Oct. 2012.

[17] G. Miao, "Energy-efficient uplink multi-user MIMO," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2302–2313, May 2013.

[18] W. A. Al-Hussaibi and F. Ali, "A closed-form approximation of correlated multiuser MIMO ergodic capacity with antenna selection and imperfect channel estimation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5515–5519, Jun. 2018.

[19] L. Dai, S. Sfar, and K. B. Letaief, "Optimal antenna selection based on capacity maximization for MIMO systems in correlated channels," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 563–573, Mar. 2006.

[20] Y. Zhang, C. Ji, W. Q. Malik, D. O'Brien, and D. J. Edwards, "Receive antenna selection for MIMO systems over correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4393–4399, Sep. 2009.

[21] W. A. Al-Hussaibi and F. H. Ali, "Fast receive antenna selection for spatial multiplexing MIMO over correlated Rayleigh fading channels," *Wireless Pers. Commun.*, vol. 70, no. 4, pp. 1243–1259, Jun. 2013.
[22] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.
[23] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
[24] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, Mar. 2016.
[25] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully non-orthogonal communication for massive access," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1717–1731, Apr. 2018.
[26] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.
[27] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec. 2017.
[28] Z. Yang, C. Pan, W. Xu, Y. Pan, M. Chen, and M. Elkashlan, "Power control for multi-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 927–942, Feb. 2018.
[29] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.
[30] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.
[31] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.
[32] Y. Liu, G. Pan, H. Zhang, and M. Song, "On the capacity comparison between MIMO-NOMA and MIMO-OMA," *IEEE Access*, vol. 4, pp. 2123–2129, 2016.
[33] Y. Chi, L. Liu, G. Song, C. Yuen, Y. L. Guan, and Y. Li, "Practical MIMO-NOMA: Low complexity and capacity-approaching solution," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6251–6264, Sep. 2018.
[34] H. Wang, R. Zhang, R. Song, and S.-H. Leung, "A novel power minimization precoding scheme for MIMO-NOMA uplink systems," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1106–1109, May 2018.
[35] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.
[36] D. Tweed and T. Le-Ngoc, "Dynamic resource allocation for uplink MIMO NOMA VWN with imperfect SIC," in *Proc. IEEE ICC*, Kansas City, MO, USA, May 2018, pp. 1–6.
[37] L. Liu, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving iterative LMMSE detection for MIMO-NOMA systems," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[38] W. A. Al-Hussaibi and F. H. Ali, "Extending the user capacity of MU-MIMO systems with low detection complexity and receive diversity," *Wireless Netw.*, vol. 24, no. 6, pp. 2237–2249, Aug. 2018.
[39] Q. Sun, S. Han, C. L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.
[40] J. Choi, "On the power allocation for MIMO-NOMA systems with layered transmissions," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3226–3237, May 2016.
[41] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.
[42] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.
[43] Y. Chen, L. Wang, Y. Ai, B. Jiao, and L. Hanzo, "Performance analysis of NOMA-SM in vehicle-to-vehicle massive MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2653–2666, Dec. 2017.
[44] X. Liu, Y. Liu, X. Wang, and H. Lin, "Highly efficient 3-D resource allocation techniques in 5G for NOMA-enabled massive MIMO and relaying systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2785–2797, Dec. 2017.
[45] X. Sun *et al.*, "Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5367–5381, Aug. 2018.
[46] K. Xiao, L. Gong, and M. Kadoch, "Opportunistic multicast NOMA with security concerns in a 5G massive MIMO system," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 91–95, Mar. 2018.
[47] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.
[48] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[49] D. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

WALID A. AL-HUSSAIBI (S'07–M'12–SM'17) received the B.Sc. degree in electronics and communications from the University of Basrah, Iraq, in 1991, the M.Sc. degree in electronics and communications from the University of Science and Technology, Jordan, in 2000, and the Ph.D. degree in wireless and mobile communications from the University of Sussex, U.K., in 2011. From 2001 to 2006, he was a Lecturer with the Department of Electrical Engineering, ETCB, FTE. In 2012, he joined Southern Technical University, Iraq, as a Faculty Member, where he is currently an Assistant Professor with the Department of ET, BTI. His research interests include massive MIMO, multiuser MIMO-NOMA, wireless communications, chaotic communications, channel modeling, capacity and performance evaluation, new modulation schemes, and multiple access techniques for future wireless systems.

FALAH H. ALI (M'00–SM'07) received the B.Sc. degree in electrical and electronics engineering and the M.Sc. degree in electronic systems from Cardiff University, in 1984 and 1986, respectively, and the Ph.D. degree in communications engineering from the University of Warwick, in 1992. From 1992 to 1994, he held a Postdoctoral research position at the University of Lancaster, developing advanced multiple access techniques for wireless communications funded by the UK research council. Between 1994 and 2008 he was a Lecturer and Senior Lecturer in electronics engineering, University of Sussex. He has also worked for the Defence sector in vehicular network enabled capabilities between 2004 and 2011. He is currently a Reader in digital communications, the Director of Communications Research Group, and the Convenor of the Master programmes on 5G mobile communications and intelligent embedded systems, at the same university. His research interests include advanced multiuser and multiantenna communication techniques, vehicular communications, wireless sensor networks, and intelligent wireless communication systems. He is a Fellow of the IET and a Chartered Engineer.